普通高中课程标准实验教科书

数学

选修 2-3

人民教育出版社 课程教材研究所
中学数学教材实验研究组 编著

人民教育出版社

B版

本册导引

我们在学习必修模块数学3的古典概型时，需要对基本事件空间中的基本事件和某一事件所包含的基本事件进行计数．细心的同学可能已经发现，那里涉及的计数几乎都可以扳着指头数数基本事件的数目来完成．随着概率学习的深入，问题变得越来越复杂，扳着指头数慢慢就行不通了，需要学习计数原理、计数方法(排列与组合)来帮助我们计数．

应用计数原理和组合数公式可以证明二项式定理，同学们在学习二项式定理时，会接触到我国古代数学的光辉成就——杨辉三角．

在本模块，同学们还将在必修模块数学3的概率知识基础上，学习二点分布、超几何分布、二项分布、正态分布及其均值、方差等内容，初步掌握利用离散型随机变量来刻画、分析随机现象的本领，从而能解决一些较简单的实际问题．

在本模块，我们还将和同学们一起进入丰富多彩的统计世界，探究独立性检验、回归分析等统计方法的奥秘．同学们在学习这两个方法时，应该充分体会统计方法的直观合理性和应用广泛性，并且要亲自动手，收集数据，开展一两次统计分析活动．

今天，假如我们回顾一下同学们的学习历程，在义务教育阶段数学课程的三个学段，在高中数学课程的必修模块中，都学习过概率与统计的知识．此外，概率思想、统计信息实际上已经渗透到了日常生活的方方面面，不断地被人们所熟知．我们深信，在这样的基础上，同学们是一定能顺利完成本模块的学习任务的．

普通高中课程标准实验教科书

数学

选修 2-3

B版

人民教育出版社　课程教材研究所
中 学 数 学 教 材 实 验 研 究 组　编著

*

出版发行

（北京市海淀区中关村南大街 17 号院 1 号楼　邮编：100081）

网址：http://www.pep.com.cn

××××印刷厂印装　全国新华书店经销

*

开本：890 毫米×1 240 毫米　1/16　印张：6.75　字数：150 000

2007 年 4 月第 2 版　　年　月第　次印刷

印数：00 001～000 000 册

ISBN 978-7-107-18754-4 / G・11844（课）　定价：　　元

著作权所有・请勿擅用本书制作各类出版物・违者必究

如发现印、装质量问题，影响阅读，请与出版社联系调换。

精品教学网www.itvb.net

全力打造全国最新最全的免费视频教学网站，现有内容已经覆盖学前，小学，初中高中，大学，职业等各学段欢迎各位爱学人士前来学习交流。

QQ309000116

目　录

主　　编　　高存明

本册主编　　高尚华
编　　者　　高尚华　邵光砚　王旭刚
　　　　　　杨　静
责任编辑　　王旭刚
美术编辑　　李宏庆　王　喆
封面设计　　李宏庆

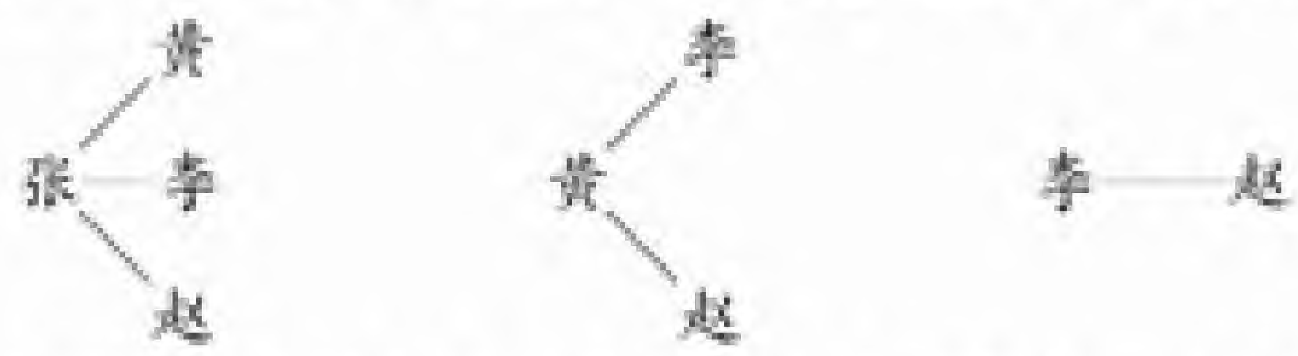
第一章　计数原理

张、黄、李、赵四位好朋友约定在寒假中要互寄贺年片一张，他们一共寄了多少张贺年片？

问题跟顺序是有关的．例如，对张、黄两人来说，张要寄给黄一张贺年片，黄要寄给张一张贺年片，这是从 4 个不同元素中取出 2 个元素的排列问题．如下图所示，一共要寄 12 张贺年片．

如果他们约定在寒假中每两人通话一次，以祝贺新年，那么，他们的通话次数一共有多少呢？

问题跟顺序无关．这句话的意思是，如果对张、黄两人来说，通话一次即可，这是从 4 个不同元素中任取 2 个元素的组合问题．如下图所示，一共通话 6 次．



当然，很多计数问题要比上述问题复杂，这也就是本章要解决的主要问题．本章将要学习两个计数原理，排列、组合的概念，排列数公式、组合数公式以及运用它们来解决一些简单的实际问题．

本章还要讨论二项式定理，同学们要了解其证明和一些简单的应用问题，这里还会涉及我国古代数学的光辉成就——杨辉三角．

1.1 基本计数原理

排列、组合的计算常常要求数一下所有可能出现的情况的个数．当问题很简单时，可以用一个一个地去数的办法解决，但对于较复杂的问题，可就数不过来了．为此我们需要掌握两个基本的计数原理：分类加法计数原理与分步乘法计数原理．

先考虑这样一个问题：从甲地到乙地，可以乘火车，也可以乘汽车，还可以乘轮船．假定火车每日 1 班，汽车每日 3 班，轮船每日 2 班，那么一天中从甲地到乙地有多少种不同的走法呢？

从甲地到乙地，可乘坐三类交通工具：火车、汽车或轮船，每类交通工具又各有若干班次．显然，选择其中任何一个班次都可以从甲地到达乙地，因此，一天中不同的走法有

$$1+3+2=6$$

种．

把上述例子推广到一般情况，就是分类加法计数原理．

分类加法计数原理　做一件事，完成它有 n 类办法，在第一类办法中有 m_1 种不同的方法，在第二类办法中有 m_2 种不同的方法……在第 n 类办法中有 m_n 种不同的方法．那么完成这件事共有

$$N=m_1+m_2+\cdots+m_n$$

种不同的方法．

> 请举出用分类形式完成工作的一个实例．

再考虑另一种形式的问题：某中学的阅览室有 50 本不同的科技书，80 本不同的文艺书．王华同学想借 1 本科技书和 1 本文艺书，共有多少种不同的借法？

借科技书和文艺书各一本，可以分成两个步骤完成：第一步借 1 本科技书，有 50 种借法；第二步借 1 本文艺书，有 80 种借法．对于 50 种科技书借法中的每一种来说，又各有 80 种文艺书借法，所以他共有

$$50\times80=4\ 000$$

种借 1 本科技书和 1 本文艺书的方法．

一般地说，有下面的分步乘法计数原理．

分步乘法计数原理　做一件事，完成它需要分成 n 个步骤，做第一个步骤有 m_1 种不同的方法，做第二个步骤有 m_2 种不同的方法……做第 n 个步骤有 m_n 种不同的方法．那么完成这件事共有

$$N=m_1\times m_2\times\cdots\times m_n$$

种不同的方法．

> 请举出用分步形式完成工作的一个实例．

> 两个基本计数原理有什么不同？

以上两个基本计数原理是解决计数问题最基本的理论依据．它们分别给出了用两种不同方式（分类和分步）完成一件事的方法总数的不同计算方法．

在“分类”问题中，各类方法中的任何一种都可以把这件事做完；在“分步”问题中，每一个步骤中的任何一种方法都不能把这件事做完，只有把各个步骤依次全部完成，才能把这件事做完.

例 1　一个三层书架的上层放有 5 本不同的数学书，中层放有 3 本不同的语文书，下层放有 2 本不同的英语书：

(1) 从书架上任取一本书，有多少种不同的取法？

(2) 从书架上任取三本书，其中数学书、语文书、英语书各一本，有多少种不同的取法？

解：(1) 从书架上任取一本书，有三类办法：

第一类办法　从书架上层任取一本数学书，有 5 种不同的方法；

第二类办法　从书架中层任取一本语文书，有 3 种不同的方法；

第三类办法　从书架下层任取一本英语书，有 2 种不同的方法.

只要在书架上任意取出一本书，任务即完成，由分类加法计数原理，不同的取法共有

$$N=5+3+2=10\text{（种）}.$$

(2) 从书架上任取三本书，其中数学书、语文书、英语书各一本，可以分成三个步骤完成：

第一步　从书架上层任取一本数学书，有 5 种不同的方法；

第二步　从书架中层任取一本语文书，有 3 种不同的方法；

第三步　从书架下层任取一本英语书，有 2 种不同的方法.

由分步乘法计数原理，不同的取法共有

$$N=5\times3\times2=30\text{（种）}.$$

所以从书架上任取三本书，其中数学书、语文书、英语书各一本，共有 30 种不同的取法.

例 2　用 0，1，2，3，4 这五个数字可以组成多少个无重复数字的：

(1) 银行存折的四位密码？

(2) 四位数？

(3) 四位奇数？

分析　(1) 可以分步选取数字，作四位密码的四个位置上的数字，且所取数字不能重复；

(2) 可以分步选取数字，分别做出千位数字、百位数字、十位数字和个位数字，且所取数字不能重复. 与(1)的不同之处是千位数字不能取 0；

(3) 四位奇数的个位只能是 1 或 3，因此符合条件的四位奇数可以分为个位数字是 1 和个位数字是 3 的两类，每一类中再分步. 要注意千位数字不能取 0，且所取数字不能重复.

解：(1) 完成“组成无重复数字的四位密码”这件事，可以分四个步骤：

第一步　选取左边第一个位置上的数字，有 5 种选取方法；

第二步　选取左边第二个位置上的数字，有 4 种选取方法；

第三步　选取左边第三个位置上的数字，有 3 种选取方法；

第四步　选取左边第四个位置上的数字，有 2 种选取方法.

由分步乘法计数原理，可组成不同的四位密码共有

$$N=5\times4\times3\times2=120\ (个).$$

（2）完成“组成无重复数字的四位数”这件事，可以分四个步骤：

第一步　从 1，2，3，4 中选取一个数字做千位数字，有 4 种不同的选取方法；

第二步　从 1，2，3，4 中剩余的三个数字和 0 共四个数字中选取一个数字做百位数字，有 4 种不同的选取方法；

第三步　从剩余的三个数字中选取一个数字做十位数字，有 3 种不同的选取方法；

第四步　从剩余的两个数字中选取一个数字做个位数字，有 2 种不同的选取方法.

由分步乘法计数原理，可组成不同的四位数共有

$$N=4\times4\times3\times2=96\ (个).$$

（3）完成“组成无重复数字的四位奇数”这件事，有两类办法：

第一类办法　四位奇数的个位取数字为 1，这件事分三个步骤完成：

第一步　从 2，3，4 中选取一个数字做千位数字，有 3 种不同的选取方法；

第二步　从 2，3，4 中剩余的两个数字与 0 共三个数字中选取一个数字做百位数字，有 3 种不同的选取方法；

第三步　从剩余的两个数字中，选取一个数字做十位数字，有 2 种不同的选取方法.

由分步乘法计数原理，第一类中的四位奇数共有

$$N_1=3\times3\times2=18\ (个).$$

第二类办法　四位奇数的个位取数字为 3，这件事分三个步骤完成：

第一步　从 1，2，4 中选取一个数字做千位数字，有 3 种不同的选取方法；

第二步　从 1，2，4 中剩余的两个数字和 0 共三个数字中选取一个数字做百位数字，有 3 种不同的选取方法；

第三步　从剩余的两个数字中，选取一个数字做十位数字，有 2 种不同的选取方法.

由分步乘法计数原理，第二类中的四位奇数共有

$$N_2=3\times3\times2=18\ (个).$$

最后，由分类加法计数原理，符合条件的四位奇数共有

$$N=N_1+N_2=18+18=36\ (个).$$

例 3　我们把壹元硬币有国徽的一面叫做正面，有币值的一面叫做反面. 现依次抛出 5 枚壹元硬币，按照抛出的顺序得到一个由 5 个“正”或“反”组成的序列，如“正、反、反、反、正”. 问：一共可以得到多少个不同的这样的序列？

例 3 与例 1、例 2 有什么不同？

解：分五个步骤完成这件事，每个步骤都有“正”或“反”两种不同的情况，由分步乘法计数原理，得

$$N=2\times2\times2\times2\times2=2^5=32.$$

所以一共可以得到 32 个不同的序列.

练习A

1. 填空：

(1) 一件工作可以用两种方法完成. 有 5 个人会用第一种方法完成，另有 4 个人会用第二种方法完成. 从这 9 个人中选出一个人来完成这件工作，不同的选法共有______种；

(2) 一个科技小组中有 3 名女同学，5 名男同学. 从中任选一名同学参加学科竞赛，共有不同的选派方法______种；若从中任选一名女同学和一名男同学参加学科竞赛，共有不同的选派方法______种.

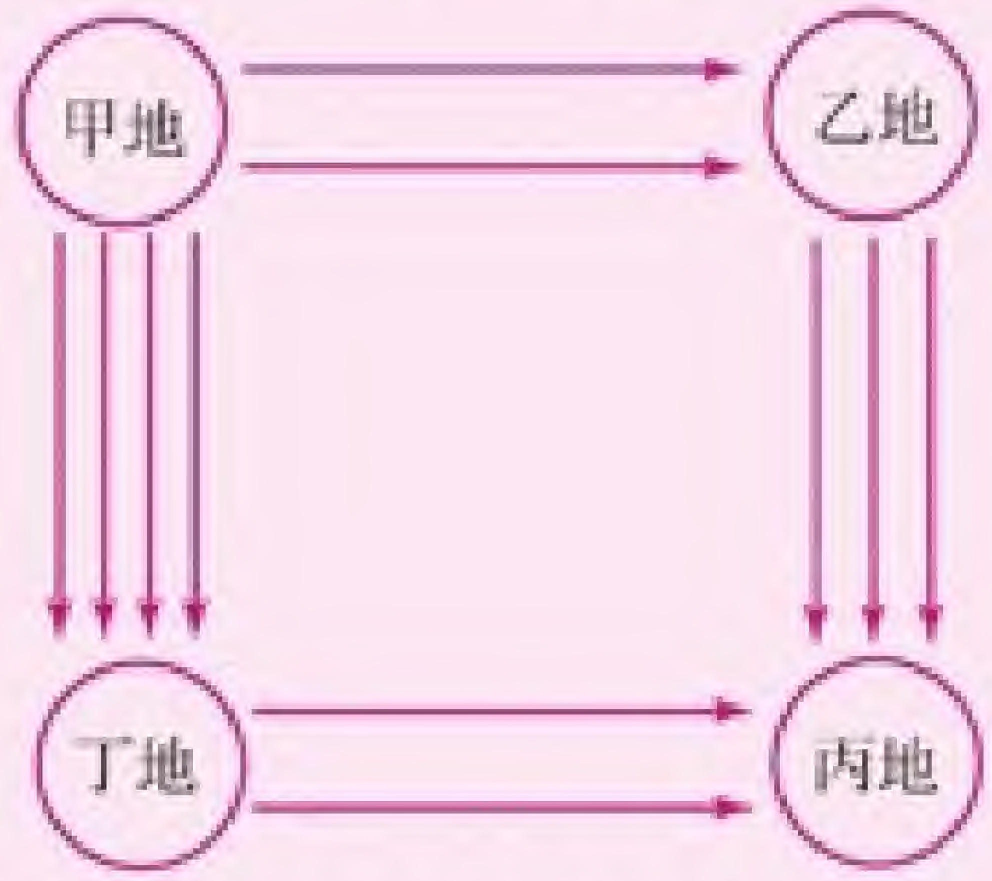
2. 一名学生做除法游戏. 在一个红口袋中装着 20 张分别标有数 1，2，…，20 的红卡片，从中任意抽取一张，把卡片上的数作为被除数；在另一个黄口袋中装着 10 张分别标有数 1，2，…，10 的黄卡片，从中任意抽取一张，把卡片上的数作为除数. 问他一共可以列出多少个不同的除法式子？

3. 从一个小组的 6 名学生中产生一名组长，一名学生代表，在下列条件下各有多少种不同的选法？

(1) 不允许兼职；　　(2) 允许兼职.

练习B

1. 如图，从甲地到乙地有 2 条路可通，从乙地到丙地有 3 条路可通；从甲地到丁地有 4 条路可通，从丁地到丙地有 2 条路可通. 问从甲地到丙地共有多少种不同的走法？



2. 由数字 0，1，2，3 这四个数字，可组成多少个：

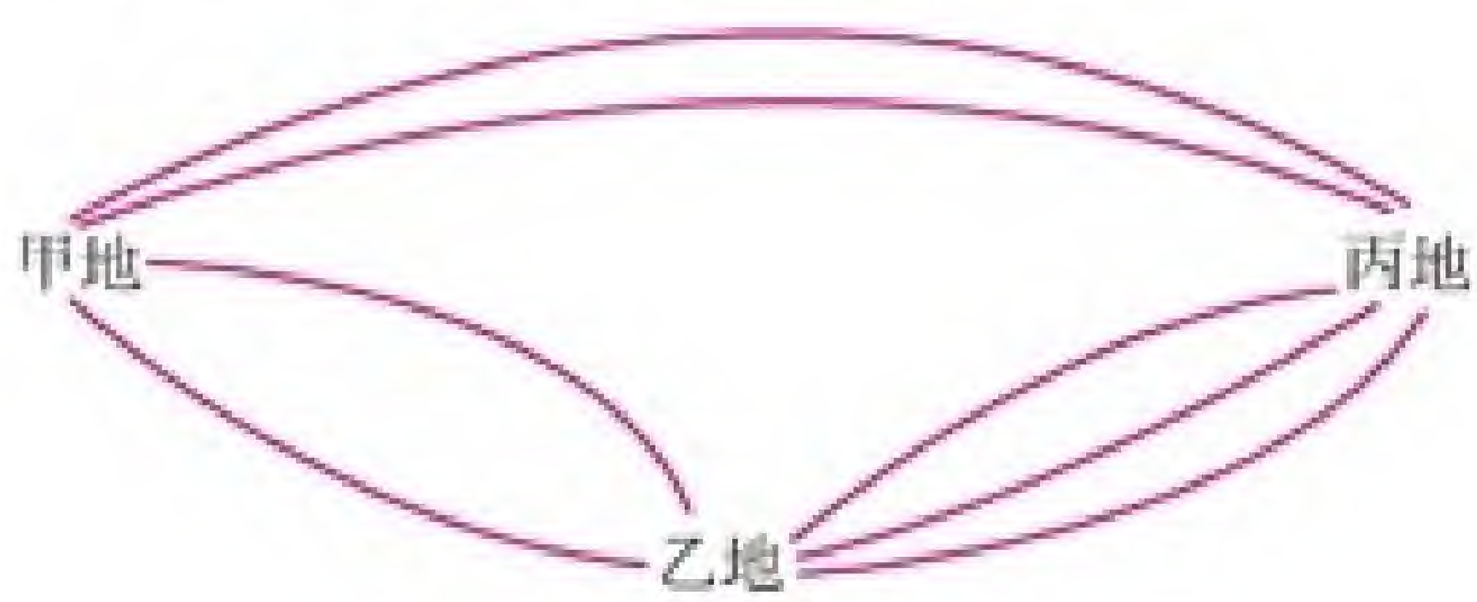
(1) 无重复数字的三位数？

(2) 可以有重复数字的三位数？

(3) 无重复数字的三位偶数？

习题 1-1 A

1. 有不同的红球 8 个，不同的白球 7 个：
   (1) 从中任意取出一个球，有多少种不同的取法？
   (2) 从中任意取出两个不同颜色的球，有多少种不同的取法？
2. 现有高一年级学生代表 3 名，高二年级学生代表 5 名，高三年级学生代表 2 名：
   (1) 从中任选一人担任校学生会主席，共有多少种不同的选法？
   (2) 从每个年级的代表中任选一人，由选出的三个人组成校学生会主席团，共有多少种不同的选法？
   (3) 从高一年级和高二年级的学生代表中各选一人，与高三年级两名学生代表，共四人组成校学生会主席团，共有多少种不同的选法？
3. 一个城市的某电话局管辖范围内的电话号码由八位数字组成，其中前四位数字是统一的，后四位数字都是 0 到 9 这十个数中的一个数字，那么不同的电话号码最多有多少个？
4. 如图，从甲地到乙地有 2 条陆路可走，从乙地到丙地有 3 条陆路可走，又从甲地不经过乙地直接到达丙地有 2 条水路可走：
   (1) 从甲地经过乙地到丙地有多少种不同的走法？
   (2) 从甲地到丙地共有多少种不同的走法？



(第 4 题)

习题 1-1 B

1. 已知集合 $M=\{1, 2, 3\}$，$N=\{2, 3, 4, 5\}$：
   (1) 任取一个奇数 n，$n\in M\cup N$，共有多少种不同的取法？
   (2) 设点 $Q(x, y)$，$x\in M$，$y\in N$，问可以表示多少个不同的点？
   (3) 在(2)中，有多少个点 $Q(x, y)$ 不在直线 $y=x$ 上？
2. 有三项体育运动项目，每个项目均设冠军和亚军各一名奖项：
   (1) 学生甲参加了这三个运动项目，但只获得一个奖项，学生甲获奖的不同情况有多少种？
   (2) 有 4 名学生参加了这三个运动项目，若一个学生可以获得多项冠军，那么各项冠军获得者的不同情况有多少种？

3. 用 0，1，…，9 十个数字，可以组成多少个：

（1）三位数？

（2）无重复数字的三位数？

（3）小于 500 的无重复数字的三位数？

（4）小于 500，且末位数字是 8 或 9 的无重复数字的三位数？

（5）小于 100 的无重复数字的自然数？

1.2 排列与组合

1.2.1 排 列

我们看下面的问题：

有红球、黄球、白球各一个，现从这三个小球中任取两个，分别放入甲、乙盒子里，有多少种不同的放法？

完成上述这件事，需要分成两个步骤：

第一步　从三个小球中任取一个放入甲盒子中，有 3 种不同的方法；

第二步　从剩下两个小球中任取一个放入乙盒子中，有 2 种不同的方法.

根据分步乘法计数原理，不同的放法共有 $3\times2=6$ 种，如图 1-1 所示：

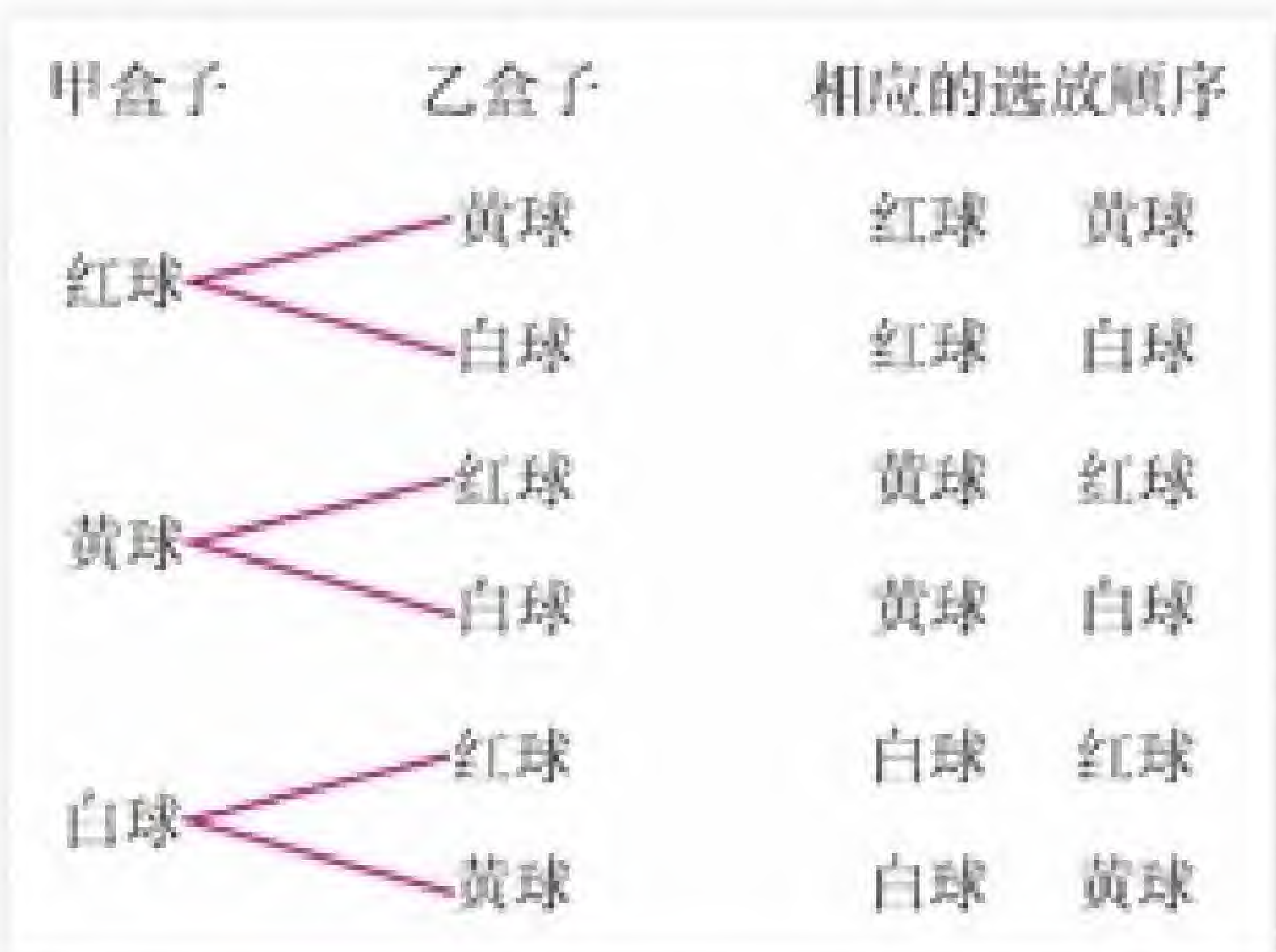


图 1-1

我们称写出所有排列的图示(图1-1)为“树形图”.

我们把被取的对象(如上面问题中的三个小球中的任何一个)叫做**元素**. 于是上述问题就抽象为：从 3 个不同元素中，任取 2 个分别占据两个位置中的一个位置，其中，选定的“位置”也可以理解成已知的“顺序”. 取出的元素占据了选定的位置后，就得到了取出元素按已知顺序排成的一列，如“红球、黄球”，我们称它为该问题的一个排列，也就是完成这件事的一种方法. 上述问题共有 6 个排列：“红球、黄球”“红球、白球”“黄球、红球”“黄球、白球”“白球、红球”“白球、黄球”，因此完成选放小球这件事，共有 6 种不同的方法.

一般地，从 n 个不同元素中任取 $m(m\leqslant n)$ 个元素，按照一定的顺序排成一列，叫做从 n 个不同元素中取出 m 个元素的**一个排列**.

根据一个排列的定义，两个排列相同的含义为：组成排列的元素相同，并且元素的排列顺序也相同.

注　一个排列就是完成一件事的一种方法；不同的排列就是完成一件事的不同方法.

从 n 个不同元素中取出 $m(m\leqslant n)$ 个元素的所有排列的个数，叫做从 n 个不同元素中取出 m 个元素的**排列数**，用符号 A_n^m（A 是英文 Arrangement(排列)的第一个字母）表示.

现在我们来研究计算排列数的公式.

一般情况下，求 A_n^m 的值，可以分成 m 个步骤完成：

第一步　从 n 个不同元素中任取一个占据第一个位置，有 n 种不同的方法；

第二步　从余下的 $(n-1)$ 个元素中任取一个占据第二个位置，有 $(n-1)$ 种不同的方法；

第三步　从余下的 $(n-2)$ 个元素中任取一个占据第三个位置，有 $(n-2)$ 种不同的方法；

> 注　所取元素不允许重复.

依此类推……

第 m 步是从前一步余下的 $[n-(m-1)]$ 个元素中任取一个占据第 m 个位置，有 $(n-m+1)$ 种不同的方法.

根据分步乘法计数原理，得到公式

$$\boxed{A_n^m=n(n-1)(n-2)\cdots(n-m+1).}$$

这里 n，$m\in \mathbf{N}_+$，并且 $m\leqslant n$. 这个公式叫做**排列数公式**. 公式右边是 m 个由大到小排列的连续正整数之积，其中最大的因数是 n，最小因数是 $(n-m+1)$.

例 1　计算从 a，b，c 这 3 个元素中，取出 3 个元素的排列数，并写出所有的排列.

解：从 a，b，c 这 3 个元素中，任取 3 个元素的排列数为

$$A_3^3=3\times2\times1=6\ (\text{个}).$$

作树形图(图 1-2).

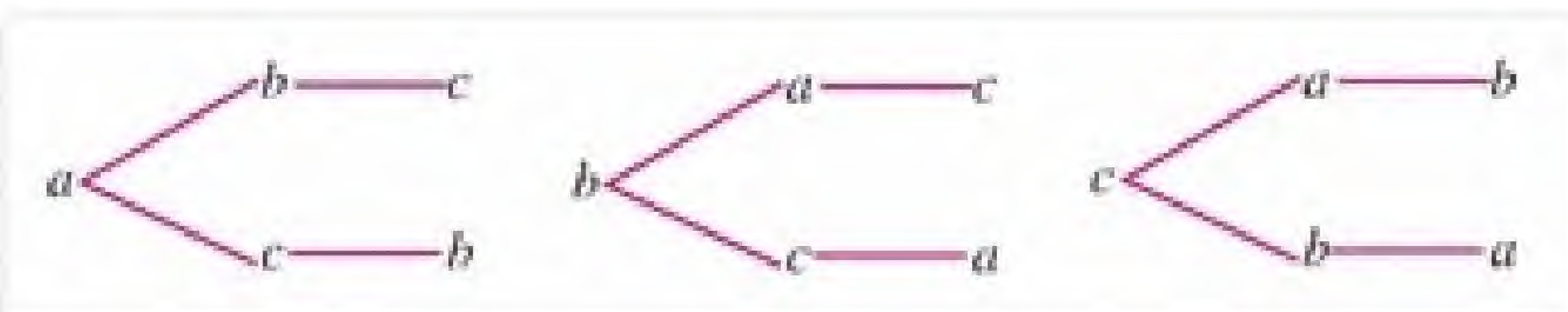


图 1-2

由图写出所有的排列为：abc，acb，bac，bca，cab，cba.

例 1 中的排列是将 a，b，c 这 3 个元素全部取出的排列问题. 其中一个排列 bac 叫做这 3 个元素的一个全排列.

一般地，n 个不同元素全部取出的一个排列，叫做 n 个不同元素的一个**全排列**. 这时在排列数公式中 $m=n$，则有

$$A_n^n=n\cdot(n-1)\cdot(n-2)\cdot\cdots\cdot3\cdot2\cdot1.$$

这个公式的右边是由 1 到 n，连续 n 个正整数的乘积. 我们把正整数由 1 到 n 的连乘积，叫做 n 的阶乘，用 $n!$ 表示. 所以 n 个不同元素的全排列数公式可以写成

$$\boxed{A_n^n=n!.}$$

我们可以将排列数公式进行下面的变形，

$$\begin{aligned}A_n^m&=n(n-1)(n-2)\cdots(n-m+1)\\&=\frac{n\cdot(n-1)\cdot(n-2)\cdot\cdots\cdot(n-m+1)\cdot(n-m)\cdot\cdots\cdot3\cdot2\cdot1}{(n-m)\cdot(n-m-1)\cdot\cdots\cdot3\cdot2\cdot1}\\&=\frac{n!}{(n-m)!}.\end{aligned}$$

因此，排列数公式还有下面的另一种形式

$$\boxed{A_n^m=\frac{n!}{(n-m)!}.}$$

排列数公式的两种不同形式，在应用中应该怎样选择？

由于 $A_n^n=n!$，为了使上面的公式在 $m=n$ 时也能成立，我们规定

$$0!=1.$$

例 2 求证：$A_n^m+mA_n^{m-1}=A_{n+1}^m$.

证明：$$\begin{aligned}A_n^m+mA_n^{m-1}&=\frac{n!}{(n-m)!}+m\frac{n!}{(n-m+1)!}\\&=\frac{n!(n-m+1)+n!\cdot m}{(n-m+1)!}\\&=\frac{n!(n-m+1+m)}{(n-m+1)!}\\&=\frac{(n+1)!}{[(n+1)-m]!}=A_{n+1}^m.\end{aligned}$$

所以 $A_n^m+mA_n^{m-1}=A_{n+1}^m$.

你能用计数原理直接解释例 2 中的等式吗？

例 3 某年全国足球中超联赛共有 12 个队参加，每队都要与其他各队在主客场分别比赛一次，共进行多少场比赛？

解：将参加比赛的 12 个队看作 12 个元素，每一场比赛即为从 12 个不同元素中任取 2 个元素的一个排列(设排在前面的队为主场比赛)．总共比赛的场次，就是从 12 个不同元素中任取 2 个元素的排列数

$$A_{12}^2=12\times11=132.$$

例 4 （1）有 3 名大学毕业生，到 5 个招聘雇员的公司应聘，若每个公司至多招聘一名新雇员，且 3 名大学毕业生全部被聘用，若不允许兼职，共有多少种不同的招聘方案？

（2）有 5 名大学毕业生，到 3 个招聘雇员的公司应聘，每个公司只招聘一名新雇员，并且不允许兼职，现假定这3 个公司都完成了招聘工作，问共有多少种不同的招聘方案？

解：（1）将 5 个招聘雇员的公司看作 5 个不同的元素，3 名大学生看作 3 个位置，则本题即为从 5 个不同元素中任取 3 个元素的排列问题．所以不同的招聘方案共有

$$A_5^3=5\times4\times3=60\text{（种）}.$$

（2）将 5 名大学毕业生看作 5 个不同的元素，3 个招聘雇员的公司看作 3 个位置，则本题仍为从 5 个不同元素中任取 3 个元素的排列问题．所以不同的招聘方案共有

$$A_5^3=5\times4\times3=60\text{（种）}.$$

例 5 某信号兵用红、黄、蓝三面旗从上到下挂在竖直的旗杆上表示信号，每次可以

挂一面、两面或三面，并且不同的顺序表示不同的信号，一共可以表示多少种不同的信号?

解：表示信号这件事，可以分为三类：

第一类 挂一面旗表示信号，是从 3 个不同元素中任取 1 个元素的排列，共有 A_3^1 种不同的方法；

第二类 挂两面旗表示信号，是从 3 个不同元素中任取 2 个元素的排列，共有 A_3^2 种方法；

第三类 挂三面旗表示信号，是 3 个元素的全排列，共有 A_3^3 种方法.

由分类加法计数原理，可以表示的信号共有

$$A_3^1+A_3^2+A_3^3=3+3\times2+3\times2\times1=3+6+6=15\ (种).$$

例 6 用 0 到 9 这十个数字可以组成多少个没有重复数字的：

(1) 三位数?

(2) 四位偶数?

解：(1) **解法一**

组成没有重复数字的三位数这件事，可以分成两个步骤完成：

第一步 从 1 到 9 这九个数字中，任取一个作为百位数字，有 A_9^1 种方法；

第二步 从 1 到 9 这些数字剩余的八个数字与 0 共九个数字中，任取两个数字依次作为十位数字和个位数字，有 A_9^2 种方法.

由分步乘法计数原理，可以组成没有重复数字的三位数共有

$$A_9^1\cdot A_9^2=9\times9\times8=648\ (个).$$

解法二

从 0 到 9 这十个数字中，任取三个数字的排列数为 A_{10}^3，其中以 0 作首位的排列数为 A_9^2. 所以可以组成没有重复数字的三位数共有

$$A_{10}^3-A_9^2=10\times9\times8-9\times8=720-72=648\ (个).$$

解法二的解题过程是：先不考虑题目中的限制条件，求出所有的排列数，然后从中减去不符合条件的排列数，就得到了所求的排列数.

> 解法二是一种用排除法思考问题的方法.

(2) 组成没有重复数字的四位偶数这件事，可以分成两类：

第一类 个位数字是 0 的四位偶数，有 A_9^3 个；

第二类 个位数字是 2，4，6，8 这四个数字之一的四位偶数，要分三个步骤完成：

第一步 从 2，4，6，8 中任取一个数字作为个位数字，有 A_4^1 种方法；

第二步 从除 0 以外的其余八个数字中，任取一个数字作为千位数字，有 A_8^1 种方法；

第三步 从剩余的八个数字中，任取两个数字作为百位和十位上的数字，有 A_8^2 种方法.

由分步乘法计数原理，第二类中的四位偶数有 $A_4^1\cdot A_8^1\cdot A_8^2$ 个.

最后，由分类加法计数原理，符合条件的四位偶数共有

$$A_9^3+A_4^1\cdot A_8^1\cdot A_8^2=504+1\ 792=2\ 296\ (个).$$

此题在求解时，还可以分成另外两类：

第一类　从1，3，5，7，9中任取一个数字作为千位数字，从0，2，4，6，8中任取一个数字作为个位数字的四位偶数，有

$$A_5^1\cdot A_5^1\cdot A_8^2\ (个).$$

第二类　从2，4，6，8中任取一个数字作为千位数字，从0，2，4，6，8中剩余的四个数字中，任取一个数字作为个位数字的四位偶数，有

$$A_4^1\cdot A_4^1\cdot A_8^2\ (个).$$

由分类加法计数原理，符合条件的四位偶数共有

$$A_5^1\cdot A_5^1\cdot A_8^2+A_4^1\cdot A_4^1\cdot A_8^2=(25+16)\times 56=2\ 296\ (个).$$

例7　有6个人排成一排：

(1) 甲和乙两人相邻的排法有多少种？

(2) 甲、乙、丙三人两两不相邻的排法有多少种？

解：(1) 甲和乙两人相邻的排列，可以分成两步完成：

第一步　将甲乙两人当作一个元素与其余四个人共5个元素排列，有 A_5^5 种方法；

第二步　将甲乙排列，有 A_2^2 种方法.

由分步乘法计数原理，共有不同的排法

$$A_5^5\cdot A_2^2=5!\cdot 2!=240\ (种).$$

(2) 甲、乙、丙三人两两不相邻的排列，分成两步完成：

第一步　除甲、乙、丙三人之外的三个人排列，有 A_3^3 种方法；

第二步　从上述排列的四个空当中任取三个空当，分别插入甲、乙、丙三人，每空一人，有 A_4^3 种方法.

由分步乘法计数原理，甲、乙、丙三人两两不相邻的排法共有

$$A_3^3\cdot A_4^3=3!\times 4\times 3\times 2=144\ (种).$$

同学们想一想：例7(2)求解时，若采用排除法，从 A_6^6 中减去不符合条件的排列数，什么样的排法不符合题中的条件呢？请你用这种方法求解后，再与例题中的解法比较，你觉得用哪种方法解更为简便？为什么？

求解排列问题时，正确地理解题意是最关键的一步，要善于把题目中的文字语言翻译成排列的相关术语. 正确运用分类加法计数原理和分步乘法计数原理是十分重要的. 分类时，要注意各类之间不重复，不遗漏. 分步时，要注意依次作完各个步骤后，事情才能完成. 如果不符合条件的情况较少时，也可以采用排除法.

一道题目，常常有多种解法，经过这种一题多解的练习，帮助同学们学会从不同的角度分析问题、解决问题.

练习A

1. 写出：

(1) 北京、上海、广州三个城市之间的所有直达航线的始发站与到达站不同的飞机票；

(2) 由1，2，3，4这四个数字组成的没有重复数字的所有四位数.

2. 计算：

(1) A_5^3；　(2) A_{15}^4；　(3) A_{100}^2；　(4) A_7^7；　(5) $A_6^4-2A_6^2$；

(6) $\dfrac{A_{12}^8}{A_{12}^7}$；　(7) $A_4^1+A_4^2+A_4^3+A_4^4$.

3. 计算2～8的阶乘数，并填入表中：

n	2	3	4	5	6	7	8
$n!$							

4. 求证：

(1) $n!=\dfrac{(n+1)!}{n+1}$；　　(2) $A_8^8-8A_7^7+7A_6^6=A_7^7$.

5. 已知 $\dfrac{A_n^7-A_n^5}{A_n^5}=89$，求 n 的值.

6. 从4种不同的蔬菜品种中选出3种，分别种植在不同土质的3块土地上进行试验，有多少种不同的种植方法？

7. 从参加乒乓球团体赛的5名运动员中选出3名参加某场比赛，每名运动员比赛一局，有多少不同的方法排定他们的出场顺序？

8. 由1，2，3，4，5，6这六个数字可组成多少个：

(1) 三位数？

(2) 没有重复数字的三位数？

(3) 没有重复数字的末位数字是5的三位数？

练习B

1. 计算下列各排列数：

(1) 从 a，b，c，d，e 中取出4个元素的排列中，a 不在首位的所有排列；

(2) 从 a，b，c，d，e 中取出4个元素的排列中，a 不在首位且 b 不在末位的所有排列.

2. 用 0，1，2，3，4，5 可组成多少个：

(1) 没有重复数字的四位数？

(2) 没有重复数字的能被 5 整除的四位数？

(3) 比 2 000 大且没有重复数字的自然数？

3. (1) 将 2 封信投入 4 个邮箱，每个邮箱最多投一封，有多少种不同的投法？

(2) 将 2 封信随意投入 4 个邮箱，有多少种不同的投法？

4. 将 2 个男生和 4 个女生排成一排：

(1) 男生排在中间的排法有多少种？

(2) 男生不在头尾的排法有多少种？

(3) 男生不相邻的排法有多少种？

(4) 男生不相邻且不在头尾的排法有多少种？

(5) 2 个男生都不与女生甲相邻的排法有多少种？

5. 四对夫妇坐成一排照相：

(1) 每对夫妇都不能隔开的排法有多少种？

(2) 每对夫妇都不能隔开，且同性别的人不能相邻的排法有多少种？

1.2.2 组 合

1. 组合及组合数公式

看下面的问题：

有红球、黄球、白球各一个，从这三个小球中，任意取出两个小球，共有多少种不同的取法？

与排列一节提出的问题的不同之处是，取出的两个小球不再放入盒子中. 我们可以一把抓出两个小球，则“从三个不同的小球中，任意取出两个小球”这件事便告完成，即取出的两个小球并无顺序关系. 上述问题中要完成的事可以抽象为：从 3 个不同元素中任取 2 个元素，不管顺序并成一组，求一共可以组成多少组？这就是本节所要研究的组合问题. 组合问题与排列问题的不同之处是：组合与取出元素的顺序无关，而排列与取出元素的顺序有关.

组合问题与排列问题最根本的不同点是什么？

一般地，从 n 个不同元素中，任意取出 $m(m\leqslant n)$ 个元素并成一组，叫做从 n 个不同元素中任取 m 个元素的**一个组合**.

一个组合就是完成事情的一种方法. 在上面的问题中，“红球、黄球”是该组合问题的一个组合.“红球、白球”是另一个组合. 根据一个组合的定义，两个组合相同的含义

为：组成组合的元素完全相同，而不管元素的顺序如何．故“红球、黄球”和“黄球、红球”是同一个组合，但不是同一个排列．

我们可以用图 1-3 的方法，写出从红、黄、白三个小球中，任意取出两个小球的所有组合：

图 1-3

故所有组合为：“红球、黄球”“红球、白球”“黄球、白球”．

所以此题的答案为：共有 3 种不同的取法．

从 n 个不同元素中，任意取出 $m(m\leqslant n)$ 个元素的所有组合的个数，叫做从 n 个不同元素中，任意取出 m 个元素的组合数，用符号 C_n^m（C 是英文字母 Combination（组合）的第一个字母）表示．

在原有知识的基础上研究新问题，是解决数学问题常用的方法．

例如上面的问题：从 3 个不同元素中任取 2 个元素的组合数表示为 C_3^2．

现在我们从研究组合数 C_n^m 与排列数 A_n^m 的关系入手，去寻找组合数 C_n^m 的计算公式．

以“从红、黄、白三个小球中，任取两个小球”的组合与排列之间的关系为例，列出下图：

图 1-4

从图 1-4 中可以看出，每一个“从红、黄、白三个小球中，任取两个小球”的组合，都对应两个“从红、黄、白三个小球中，任取两个小球”的排列．因此有

$$A_3^2=C_3^2\cdot A_2^2.$$

这个等式表明，从 3 个不同元素中任取 2 个元素的排列，可以分两步完成：

第一步　选取元素　从 3 个不同元素中任取 2 个元素的组合，共有 C_3^2 种方法；

第二步　排位置　选出的 2 个不同元素的全排列，有 A_2^2 种方法．

一般地，从 n 个不同元素中，任取 m 个元素的排列，可以分两步完成：

第一步　选取元素　从 n 个不同元素中，任取 m 个元素的组合，有 C_n^m 种方法；

第二步　排位置　选出的 m 个不同元素的全排列，有 A_m^m 种方法．

根据分步乘法计数原理，得

$$\boxed{A_n^m=C_n^m\cdot A_m^m.}$$

这个公式不仅揭示了组合数 C_n^m 与排列数 A_n^m 之间的关系，也表明解某些排列问题时，常常分选元素和排位置两个步骤完成．

由于 A_n^m 计算公式和

$$C_n^m=\frac{A_n^m}{A_m^m},$$

得出组合数 C_n^m 计算公式为

$$C_n^m=\frac{n(n-1)(n-2)\cdots(n-m+1)}{m!}; \qquad ①$$

$$C_n^m=\frac{n!}{m!(n-m)!}. \qquad ②$$

在组合数计算公式②中，当 $m=n$ 时，由于 $0!=1$，故有

$$C_n^n=\frac{n!}{n!\cdot 0!}=1.$$

当 $m=0$ 时，组合数公式仍有意义，将 $m=0$ 代入组合数计算公式②中，得

$$C_n^0=\frac{n!}{n!\cdot 0!}=1.$$

所以

$$\boxed{C_n^0=1.}$$

对于组合数，C_n^m 应有 $m\in\mathbf{N}$，$n\in\mathbf{N}^+$，且 $m\leqslant n$.

例 1 计算：

(1) $C_7^3+C_7^4$； (2) $C_{10}^5\cdot C_{10}^0-C_{10}^{10}$.

解：(1) $C_7^3+C_7^4=\frac{7\times6\times5}{3\times2\times1}+\frac{7\times6\times5\times4}{4\times3\times2\times1}=35+35$

$=70$；

(2) $C_{10}^5\cdot C_{10}^0-C_{10}^{10}=\frac{10\times9\times8\times7\times6}{5\times4\times3\times2\times1}\times1-1=252-1$

$=251$.

例 2 平面内有 10 个点，其中任何 3 个点不共线，以其中任意 2 个点为端点的

(1) 线段有多少条？

(2) 有向线段有多少条？

> 若去掉例 2 中“任何 3 个点不共线”的条件，题目应该怎样设问？

解：(1) 所求线段的条数，即为从 10 个元素中任取 2 个元素的组合，共有

$$C_{10}^2=\frac{10\times9}{2\times1}=45\ (条),$$

即以 10 个点中的 2 个点为端点的线段共有 45 条.

(2) 所求有向线段的条数，即为从 10 个元素中任取 2 个元素的排列，共有

$$A_{10}^2=10\times9=90\ (条),$$

即以 10 个点中的 2 个点为端点的有向线段共有 90 条.

2. 组合数的两个性质

在例 1 中，我们计算得 $C_7^3=35$，$C_7^4=35$，故 $C_7^3=C_7^4$，也可以表示为

$$C_7^3=C_7^{7-3}.$$

在一般的情况下，下面的性质成立：

> 这个等式为什么会成立呢？

性质 1

$$\boxed{C_n^m=C_n^{n-m}.}$$

探索与研究

性质 1 成立的理由是 $C_n^m=\dfrac{n!}{m!(n-m)!}=\dfrac{n!}{(n-m)![n-(n-m)]!}=C_n^{n-m}$.

为了简化计算，当 $m>\dfrac{n}{2}$ 时，通常将计算 C_n^m 化为计算 C_n^{n-m}. 如

$$C_7^4=C_7^3=\frac{7\times6\times5}{3\times2\times1}=35.$$

例 3 一个口袋里装有 7 个不同白球和 1 个红球，从口袋中任取 5 个球：

(1) 共有多少种不同的取法?

(2) 其中恰有一个红球，共有多少种不同的取法?

(3) 其中不含红球，共有多少种不同的取法?

解：(1) 从口袋里的 8 个球中任取 5 个球，不同取法的种数是

$$C_8^5=C_8^3=\frac{8\times7\times6}{3\times2\times1}=56.$$

(2) 从口袋里的 8 个球中任取 5 个球，其中恰有一个红球，可以分两步完成：

第一步　从 7 个白球中任取 4 个白球，有 C_7^4 种取法；

第二步　把 1 个红球取出，有 C_1^1 种取法.

由分步乘法计数原理，不同取法的种数是

$$C_7^4\cdot C_1^1=C_7^4=C_7^3=35.$$

(3) 从口袋里任取 5 个球，其中不含红球，只需从 7 个白球中任取 5 个白球即可，不同取法的种数是

$$C_7^5=C_7^2=\frac{7\times6}{2\times1}=21.$$

从上面三道小题的答案可以得出等式

$$C_8^5=C_7^4+C_7^5.$$

> 这个等式的成立，是确有理论依据，还是由巧合造成的?

一般地，从$(n+1)$个不同元素中任取 m 个元素的组合，可以分为两类：第一类取出的 m 个元素中不含某个元素 a 的组合，只需在除去元素 a 的其余 n 个元素中任取 m 个，有 C_n^m 个；第二类取出的 m 个元素中含有某个元素 a 的组合，只需在除去元素 a 的其余 n 个元素中任取$(m-1)$个后再取出元素 a，有 C_n^{m-1} 个.

根据分类加法计数原理，得出组合数的另一个性质：

性质 2

$$\boxed{C_{n+1}^m=C_n^m+C_n^{m-1}.}$$

探索与研究

性质 2 成立的理由是
$$\begin{aligned}C_n^m+C_n^{m-1}&=\frac{n!}{m!(n-m)!}+\frac{n!}{(m-1)!(n-m+1)!}\\&=\frac{n!(n-m+1)+n!m}{m!(n-m+1)!}\\&=\frac{n!(n+1)}{m!(n-m+1)!}\\&=\frac{(n+1)!}{m![(n+1)-m]!}=C_{n+1}^m.\end{aligned}$$

组合数的这个性质，在简化计算和在某些关于组合数的式子变形中常有应用，如
$$C_9^3+C_9^2=C_{10}^3=\frac{10\times9\times8}{3\times2\times1}=120.$$

例 4 在产品质量检验时，常从产品中抽出一部分进行检查. 现在从 98 件正品和 2 件次品共 100 件产品中，任意抽出 3 件检查：

（1）共有多少种不同的抽法？

（2）恰好有一件是次品的抽法有多少种？

（3）至少有一件是次品的抽法有多少种？

（4）恰好有一件是次品，再把抽出的 3 件产品放在展台上，排成一排进行对比展览，共有多少种不同的排法？

解：（1）所求不同的抽法数，即从 100 个不同元素中任取 3 个元素的组合数，共有
$$C_{100}^3=\frac{100\times99\times98}{3\times2\times1}=161\ 700\text{（种）}.$$

（2）抽出的 3 件中恰好有一件是次品这件事，可以分两步完成：

第一步　从 2 件次品中任取 1 件，有 C_2^1 种方法；

第二步　从 98 件正品中任取 2 件，有 C_{98}^2 种方法.

根据分步乘法计数原理，不同的抽取方法共有

$$C_2^1 \cdot C_{98}^2=2\times\frac{98\times97}{2\times1}=2\times4\,753=9\,506\text{（种）}.$$

（3）解法一　抽出的 3 件中至少有一件是次品这件事，分为两类：

第一类　抽出的 3 件中有 1 件是次品的抽法，有 $C_2^1C_{98}^2$ 种；

第二类　抽出的 3 件中有 2 件是次品的抽法，有 $C_2^2C_{98}^1$ 种.

根据分类加法计数原理，不同的抽法共有

$$C_2^1 \cdot C_{98}^2+C_2^2 \cdot C_{98}^1=9\,506+98=9\,604\text{（种）}.$$

解法二　从 100 件产品中任取 3 件的抽法，有 C_{100}^3 种，其中抽出的 3 件中没有次品的抽法，有 C_{98}^3 种. 所以抽出的 3 件中至少有一件是次品的抽法，共有

$$C_{100}^3-C_{98}^3=161\,700-152\,096=9\,604\text{（种）}.$$

（4）完成题目中的事，可以分成两步：

第一步　选取产品，有 $C_2^1C_{98}^2$ 种方法；

第二步　选出的 3 个产品排列，有 A_3^3 种方法.

根据分步乘法计数原理，不同的排法共有

$$C_2^1C_{98}^2A_3^3=57\,036\text{（种）},$$

即抽出的 3 件产品中有一件次品，且排成一排的方法共有 57 036 种.

例 5　有 9 本不同的课外书，分给甲、乙、丙三名同学，求在下列条件下，各有多少种不同的分法？

（1）甲得 4 本，乙得 3 本，丙得 2 本；

（2）一人得 4 本，一人得 3 本，一人得 2 本；

（3）甲、乙、丙各得 3 本.

解：（1）甲得 4 本，乙得 3 本，丙得 2 本这件事分三步完成：

第一步　从 9 本不同的书中，任取 4 本分给甲，有 C_9^4 种方法；

第二步　从余下的 5 本书中，任取 3 本分给乙，有 C_5^3 种方法；

第三步　把剩下的 2 本书给丙，有 C_2^2 种方法.

根据分步乘法计数原理，共有不同的分法

$$C_9^4 \cdot C_5^3 \cdot C_2^2=C_9^4 \cdot C_5^3=1\,260\text{（种）}.$$

所以甲得 4 本，乙得 3 本，丙得 2 本的分法共有1 260种.

（2）一人得 4 本，一人得 3 本，一人得 2 本这件事，分两步完成：

第一步　按 4 本、3 本、2 本分成三组，有 $C_9^4C_5^3C_2^2$ 种方法；

第二步　将分成的三组书分给甲、乙、丙三个人，有 A_3^3 种方法.

根据分步乘法计数原理，共有不同的分法

$$C_9^4C_5^3C_2^2A_3^3=7\,560\text{（种）}.$$

所以一人得 4 本，一人得 3 本，一人得 2 本的分法共有 7 560 种.

（3）用与(1)相同的方法求解，得 $C_9^3 \cdot C_6^3 \cdot C_3^3=1\,680$（种）.

所以甲、乙、丙各得 3 本的分法共有 1 680 种.

例 6　某次足球赛共 12 支球队参加，分三个阶段进行：

(1) 小组赛：经抽签分成甲、乙两组，每组6队进行单循环比赛，以积分及净剩球数取前两名；

(2) 半决赛：甲组第一名与乙组第二名，乙组第一名与甲组第二名作主客场交叉淘汰赛(每两队主客场各赛一场)决出胜者；

(3) 决赛：两个胜队参加决赛一场，决出胜负.

问全部赛程共需比赛多少场？

解：(1) 小组赛中每组6队进行单循环比赛，就是6支球队的任两支球队都要比赛一次，所需比赛的场次即为从6个元素中任取2个元素的组合数，所以小组赛共要比赛

$$2C_6^2=2\times\frac{6\times5}{1\times2}=30\text{（场）}.$$

(2) 半决赛中甲组第一名与乙组第二名(或乙组第一名与甲组第二名)主客场各赛一场，所需比赛的场次即为从2个元素中任取2个元素的排列数，所以半决赛共要比赛

$$2A_2^2=2\times1\times2=4\text{（场）}.$$

(3) 决赛只需比赛1场，即可决出胜负.

所以全部赛程共需比赛

$$30+4+1=35\text{（场）}.$$

例7 设北京故宫博物院某日接待游客10 000人，如果从这些游客中任意选出10名幸运游客，一共有多少种不同的选择（保留4位有效数字）？若把10份不同的纪念品发给选出的幸运游客每人一份，又有多少种不同的选择？

解：从10 000人中选出10人的选择有$C_{10\,000}^{10}$种，用计算器计算：

按键	显示
10000 [nCr] 10 [=]	2.743355078^{33}

所以　$C_{10\,000}^{10}\approx2.743\times10^{33}$，一共约有$2.743\times10^{33}$种不同的选择.

10个人领取10份不同的纪念品的情况有A_{10}^{10}种，用计算器计算：

按键	显示
10 [SHIFT] [nAr] 10 [=]	3628800

所以　$A_{10}^{10}=3\ 628\ 800$，又有3 628 800种不同的选择.

计算A_{10}^{10}时还可通过$A_{10}^{10}=10!$计算：

按键	显示
10 [SHIFT] [X!] 10 [=]	3628800

练习A

1. 北京、上海、天津、广东四个足球队举行友谊比赛，每两个队都要比赛一场：

(1) 列出所有各场比赛的双方；

(2) 列出所有冠亚军的可能情况.

2. 写出：

(1) 从 a，b，c，d，e 五个元素中任取两个元素的所有组合；

(2) 从 a，b，c，d，e 五个元素中任取三个元素的所有组合.

3. 计算：

(1) $C_6^2+C_8^3$；　　(2) $C_{10}^6-C_8^3$；

(3) $C_{100}^{98}\div C_9^7$；　　(4) $C_{12}^5+C_{12}^6$；

(5) $C_5^0+C_5^1+C_5^2+C_5^3+C_5^4+C_5^5$；　　(6) $C_7^3+C_7^4+C_8^5+C_9^6+C_{10}^7$.

4. 从 3，5，7，11 这四个质数中任取两个相乘，可以得到多少个不相等的积？

5. 某校举行排球赛，每两个队赛一场，有 8 个队参加，共需比赛多少场？

6. 有 8 名男生和 5 名女生，从中任选 6 人：

(1) 有多少种不同的选法？

(2) 其中有 3 名女生，共有多少种不同的选法？

(3) 其中至多有 3 名女生，共有多少种不同的选法？

(4) 其中有 2 名女生、4 名男生，分别担任 6 种不同的工作，共有多少种不同的分工方法？

练习B

1. 解方程：$C_{18}^{x}=C_{18}^{3x-6}$.

2. 在 10 件产品中，有 3 件次品，从中任取 5 件：

(1) 恰有 2 件次品的抽法有多少种？

(2) 至多有 2 件次品的抽法有多少种？

(3) 至少有 1 件次品的抽法有多少种？

(4) 至少有 2 件次品，2 件正品的抽法有多少种？

3. 从 1，3，5，7，9 中任取三个数字，从 2，4，6，8 中任取两个数字，可以组成多少：

(1) 无重复数字的五位数？

(2) 万位、百位和个位数字是奇数的无重复数字的五位数？

(3) 千位和十位数字只能是奇数的无重复数字的五位数？

(4) 其和不等的加法算式？

4. 甲、乙、丙、丁、戊五名同学参加某种技术竞赛，决出了第一名到第五名的五个名次．甲、乙两人去询问成绩，组织者对甲说：“很遗憾，你和乙都未拿到冠军”；对乙说：“你当然不会是最差的”．从他的回答分析，这五个人的名次排列共有多少种不同的情况？

5. 把四个不同的小球放入三个分别标有 1～3 号的盒子中：

(1) 不许有空盒子的放法有多少种？

(2) 允许有空盒子的放法有多少种？

(3) 若把四个小球分别标上 1～4 的标号，不许有空盒子且任意一个小球都不能放入标有相同标号的盒子中，共有多少种不同的放法？

6. 将 6 名应届大学毕业生分配到 3 个公司：

(1) 3 个人分到甲公司，2 个人分到乙公司，1 个人分到丙公司，有多少种不同的分配方案？

(2) 一个公司去 3 个人，另一个公司去 2 个人，剩下的一个公司去 1 个人，有多少种不同的分配方案？

习题 1-2 A

1. 计算：

(1) $5A_5^3+4A_4^2$；　　(2) $A_4^4+A_5^2+A_4^3+A_5^5$；

(3) $\dfrac{A_8^8}{A_8^7}+A_8^2$；　　(4) $\dfrac{A_7^5-A_6^6}{7!+6!}$．

2. 求下列各式中的 n：

(1) $A_{2n}^3=10A_n^3$；　　(2) $\dfrac{A_n^5+A_n^4}{A_n^3}=4$．

3. 填空：

(1) 已知 $A_{10}^m=10\times9\times8\times7\times6\times5$，那么 $m=$______；

(2) 已知从 n 个不同元素中取出 2 个元素的排列数是 56，则 $n=$______；

(3) 已知从 n 个不同元素中取出 2 个元素的排列数等于从 $(n-4)$ 个不同元素中取出 2 个元素的排列数的 7 倍，则 $n=$______．

4. 一个火车站有 8 股岔道，停放 4 列不同的火车，有多少种不同的停放方法（假定每股岔道上只能停放一列火车）．

5. 一部记录影片在 4 个单位轮流放映，每一个单位放映一场，可有多少种不同的放映次序？

6. (1) 由数字 1，2，3，4，5，6 可以组成多少个没有重复数字的五位数？

(2) 由数字 0，1，2，3，4，5 可以组成多少个没有重复数字的五位数、五位偶数、自然数？

7. 计算：

(1) C_{15}^{3}；　　(2) C_{200}^{197}；

(3) $C_{11}^{4}-C_{9}^{3}$；　　(4) $C_{10}^{4}+C_{8}^{5}$；

(5) $C_{6}^{3}\div C_{8}^{4}$；　　(6) $C_{n+1}^{n}\cdot C_{n}^{n-1}$.

8. 圆上有 10 个点.

(1) 过每两个点可画一条弦，一共可画多少条弦？

(2) 过每 3 个点可画一个圆内接三角形，一共可画多少个圆内接三角形？

9. (1) 空间有 8 个点，没有 4 个点在同一平面内，过每 3 个点作一个平面，一共可以作多少个平面？

(2) 空间 10 个点，其中任何 4 个点不共面，以每 4 个点为顶点作一个四面体，一共可以作多少个四面体？

10. 乒乓球邀请赛先分成三个组，第一、二组各有 7 个队，第三组有 6 个队进行小组单循环赛，然后各小组的第一名共三个队分主客场进行决赛，最终决出冠亚军，一共需要比赛多少场？

11. 某班有 52 名学生，其中正、副班长各一名，现选派 5 名学生参加某种活动：

(1) 如果正、副班长必须在内，有多少种选派方法？

(2) 如果正、副班长必须有一人在内，且只能有一人在内，有多少种选派方法？

(3) 如果正、副班长都不在内，有多少种选派方法？

(4) 如果正、副班长至少有一人在内，有多少种选派方法？

12. 有 6 名女生，4 名男生，从中选出 3 名女生和 2 名男生：

(1) 组成班委会，有多少种不同的选法？

(2) 选出的 5 名学生分别担任班委会中的 5 种不同的工作，有多少种选派方法？

(3) 女生担任班长、学习委员和文娱委员，男生担任宣传委员和体育委员，有多少种选派方法？

13. (1) 平面内有两组平行线，一组有 m 条，另一组有 n 条. 这两组平行线相交，可以构成多少个平行四边形？

(2) 空间有三组平行平面，第一组有 m 个，第二组有 n 个，第三组有 l 个. 不同两组的平面都相交，且交线都不平行，可构成多少个平行六面体？

习题 1-2 B

1. 计算：$C_{2}^{2}+C_{3}^{2}+C_{4}^{2}+\cdots+C_{100}^{2}$.

2. 求证：$A_{m}^{m}+A_{m+1}^{m}+A_{m+2}^{m}+\cdots+A_{2m}^{m}=A_{2m+1}^{m}$.（提示：根据排列数和组合数间的关系，化为组合数求和.）

3. 由数字 0，1，2，3，4，5 可以组成多少个比 200 000 大且千位数字比 4 小的没有重复数字的六位数?

4. 在 3 000～7 000 之间有多少个没有重复数字的 5 的倍数?

5. 有 6 个人分成两排就座，每排 3 人：

(1) 有多少种不同的坐法?

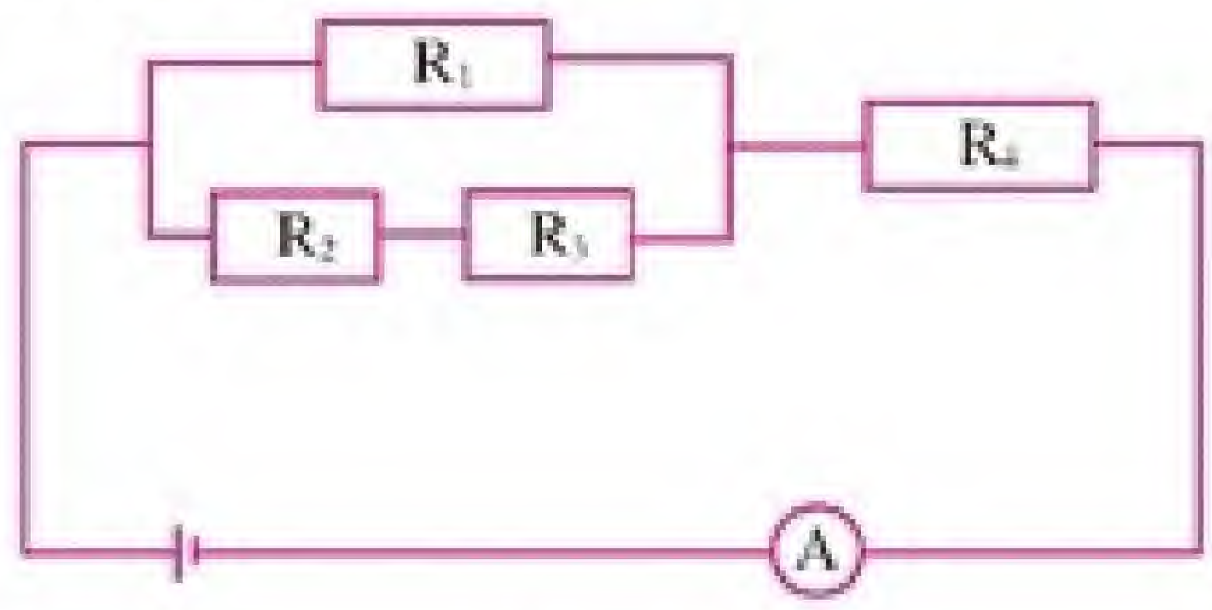
(2) 如果甲不能坐在第一排，乙不能坐在第二排，有多少种不同的坐法?

(3) 如果甲和乙必须在同一排且相邻，有多少种不同的坐法?

(4) 如果甲和乙必须在同一排且不相邻，有多少种不同的坐法?

6. 有甲、乙、丙三项任务，甲需要 2 个人承担，乙、丙各需 1 人承担．从 10 个人中选派 4 个人承担这三项任务，不同的选法有多少种?

7. 如图，电路中有 4 个电阻和 1 个电流表 A，若没有电流通过电流表 A，其原因仅因为电阻断路的可能情况共有多少种?



(第 7 题)

8. 某教师一个上午有 3 个班的课，每班一节．如果上午只能排 4 节课，并且教师不能连上 3 节课，那么这位教师上午的课表有多少种可能的排法?

9. 某种产品的加工需要经过 5 道工序：

(1) 如果其中某一工序不能放在最后加工，共可以有多少种加工顺序?

(2) 如果其中两道工序既不能放在最前，也不能放在最后，共可以有多少种加工顺序?

10. 圆周上有 n ($n>5$)个点，用线段将它们中的任意两个点相连．这些线段中任意三条在圆内都不交于一点，问：这些线段能构成多少个顶点在圆内的三角形?

11. 将 4 封信全部投入 3 个邮筒，

(1) 每个邮筒至少投一封，有多少种不同的投法?

(2) 可以随意投，有多少种不同的投法?

12. 有 6 个座位连成一排，安排 3 个人就座，恰有两个空位相邻的不同坐法共有多少种?

1.3 二项式定理

1.3.1 二项式定理

同学们在初中代数中学过乘法公式

$$(a+b)^2=a^2+2ab+b^2.$$

可以由$(a+b)^3=(a+b)^2(a+b)$，计算得

$$(a+b)^3=a^3+3a^2b+3ab^2+b^3.$$

同样，可以由$(a+b)^4=(a+b)^3(a+b)$，计算得$(a+b)^4$的展开式

$$(a+b)^4=a^4+4a^3b+6a^2b^2+4ab^3+b^4.$$

为了便于应用，需要研究$(a+b)^n$的展开式中各项是怎样组成的？

> 本节知识是初中所学乘法公式的扩展．公式是数学知识的一个重要的组成部分，应用它可以简化计算．

先以$(a+b)^4$的展开式为例进行分析，由于

$$(a+b)^4=(a+b)(a+b)(a+b)(a+b),$$

右式展开后的每一项是从每个括号里任取一个字母的乘积，因此$(a+b)^4$展开后各项都是4次式，分别为

$$a^4,\ a^3b,\ a^2b^2,\ ab^3,\ b^4.$$

以组合的观点来看，组成$(a+b)^4$的展开式中的每一项，都需要分两个步骤完成，即第一步取b；第二步取a.

a^4项：从四个因式中都不取b，而全部取a，共有$C_4^0C_4^4$个a^4，所以$(a+b)^4$的展开式的第1项为$C_4^0a^4$；

a^3b项：从四个因式中取出一个因式中的b，再从其余三个因式中全部取a，共有$C_4^1C_3^3$个a^3b，所以$(a+b)^4$的展开式的第2项为$C_4^1a^3b$；

a^2b^2项：从四个因式中取出两个因式中的b，再从其余两个因式中全部取a，共有$C_4^2C_2^2$个a^2b^2，所以$(a+b)^4$的展开式的第3项为$C_4^2a^2b^2$；

ab^3项：从四个因式中取出三个因式中的b，再从其余一个因式中取a，共有$C_4^3C_1^1$个ab^3，所以$(a+b)^4$的展开式的第4项为$C_4^3ab^3$；

b^4项：从四个因式中全部取b，而不取a，共有C_4^4个b^4，所以$(a+b)^4$的展开式的第5项为$C_4^4b^4$.

因此，$(a+b)^4=C_4^0a^4+C_4^1a^3b+C_4^2a^2b^2+C_4^3ab^3+C_4^4b^4$.

> 你会用类似的方法求$(x-2y+z)^5$展开式中x^2y^2z项的系数吗？

一般地，我们可以把$(a+b)^n$表示为

$$(a+b)^n=\underbrace{(a+b)(a+b)\cdots(a+b)}_{n个}.$$

由前面的分析可知，n 个$(a+b)$相乘，乘得的每一项，是从每个$(a+b)$里任取一个字母 a 或 b 的乘积，所以$(a+b)^n$ 的展开式中每一项都是 $a^{n-r}b^r$ 的形式($r=0$，1，…，n)．对于每一项 $a^{n-r}b^r$，是由 r 个$(a+b)$选了 b，$n-r$ 个$(a+b)$选了 a 得到的，它出现的次数(也就是这一项的系数)相当于从 n 个$(a+b)$中取 r 个 b 的组合数，即为 C_n^r，由此得到下面的公式：

$$(a+b)^n=C_n^0a^n+C_n^1a^{n-1}b+C_n^2a^{n-2}b^2+\cdots+C_n^ra^{n-r}b^r+\cdots+C_n^nb^n\ (n\in\mathbf{N}_+).$$

这个公式所表示的规律叫做**二项式定理**，等式右边的多项式叫做$(a+b)^n$ 的**二项展开式**，它一共有 $n+1$ 项，其中各项系数 C_n^r($r=0$，1，…，n)叫做展开式的**二项式系数**．展开式中的 $C_n^ra^{n-r}b^r$ 项叫做二项展开式的**通项**，通项是展开式的第 $r+1$ 项，即

$$T_{r+1}=C_n^ra^{n-r}b^r$$
$$(\text{其中 } 0\leqslant r\leqslant n,\ r\in\mathbf{N},\ n\in\mathbf{N}_+).$$

为什么不把展开式中的第 r 项 T_r 定为二项展开式的通项呢?

我们把上面的公式叫做二项展开式的**通项公式**.

在二项式定理中，如果设 $a=1$，$b=x$，则得到公式

$$(1+x)^n=1+C_n^1x+C_n^2x^2+\cdots+C_n^rx^r+\cdots+C_n^nx^n.$$

例 1　展开$\left(2\sqrt{x}-\dfrac{1}{\sqrt{x}}\right)^6$.

注　也可以将原式中的根式化为分数指数，分式化为负指数，再展开.

解：
$$\left(2\sqrt{x}-\frac{1}{\sqrt{x}}\right)^6=\left(\frac{2x-1}{\sqrt{x}}\right)^6=\frac{(2x-1)^6}{x^3}$$
$$=\frac{1}{x^3}[2x+(-1)]^6$$
$$=\frac{1}{x^3}[(2x)^6-C_6^1(2x)^5+C_6^2(2x)^4-C_6^3(2x)^3+C_6^4(2x)^2-C_6^5(2x)+C_6^6]$$
$$=\frac{1}{x^3}(64x^6-192x^5+240x^4-160x^3+60x^2-12x+1)$$
$$=64x^3-192x^2+240x-160+\frac{60}{x}-\frac{12}{x^2}+\frac{1}{x^3}.$$

例 2　求$(1+2x)^7$的展开式的第 4 项的二项式系数和系数.

注　展开式中各项的二项式系数和该项的系数是两个不同的概念，只有在特殊情况下，它们的值才相等.

解：$(1+2x)^7$的展开式的第 4 项是

$$T_4=T_{3+1}=C_7^3\cdot 1^4\cdot(2x)^3=C_7^3\cdot 2^3\cdot x^3,$$

所以展开式的第 4 项的二项式系数是 $C_7^3=35$，展开式的第 4 项的系数是 $C_7^3\cdot 2^3=280$.

例 3　求$\left(x-\dfrac{1}{x}\right)^9$展开式中含 x^3 的项，并说明它是展开式的第几项？

解：$\left(x-\dfrac{1}{x}\right)^9$展开式的通项是

$$T_{r+1}=C_9^r x^{9-r}\left(-\frac{1}{x}\right)^r=(-1)^r C_9^r x^{9-2r}.$$

依题意，有

$$9-2r=3,\ r=3.$$

所以，展开式中含 x^3 的项为

$$T_{3+1}=(-1)^3C_9^3x^3=-84x^3,$$

它是展开式的第 4 项.

练习A

1. 写出$(p+q)^7$的展开式.
2. 求$(2a+3b)^6$展开式中的第 3 项.
3. 写出$\left(\sqrt[3]{x}-\dfrac{1}{2\sqrt[3]{x}}\right)^n$展开式的通项.
4. 填空：$(x^3+2x)^7$展开式的第 4 项是______；第 4 项的二项式系数是______；第 4 项的系数是______.

练习B

1. 求$\left(a^2+\dfrac{1}{a}\right)^{15}$展开式中含 a^9 项的系数.
2. 求$\left(\sqrt{x}-\dfrac{3}{\sqrt{x}}\right)^{10}$展开式中含 x^4 的项.
3. 求$\left(x-\dfrac{1}{x}\right)^8$展开式中的常数项.

1.3.2 杨辉三角

$(a+b)^n$ 展开式的二项式系数，当 n 取正整数时可以单独列成下表的形式：

$(a+b)^1$…………………… 1 1

$(a+b)^2$………………… 1 2 1

$(a+b)^3$……………… 1 3 3 1

$(a+b)^4$…………… 1 4 6 4 1

$(a+b)^5$………… 1 5 10 10 5 1

$(a+b)^6$……… 1 6 15 20 15 6 1

……

图 1-5

上面的二项式系数表称为“杨辉三角”或“贾宪三角”．杨辉是我国宋朝数学家，他于 1261 年著《详解九章算法》，在其中详细列出了这样一张图表(图 1-5)，并且指出这个方法出于更早期贾宪的著作《黄帝九章算法细草》．在欧洲一般认为这是帕斯卡(Pascal)于 1654 年发现的，称这个图形为“帕斯卡三角”．

观察杨辉三角，可以看出二项式系数具有下面 3 条性质：

1. 每一行的两端都是 1，其余每个数都等于它“肩上”两个数的和．

性质 1 实际上反映了组合数的下列性质：

$$C_n^0=1,\ C_n^n=1,$$
$$C_{n+1}^m=C_n^{m-1}+C_n^m.$$

2. 每一行中，与首末两端“等距离”的两个数相等．

这就是说，二项展开式中，与首末两端“等距离”的两项的二项式系数相等，实际上反映了组合数的性质

$$C_n^m=C_n^{n-m}.$$

3. 如果二项式的幂指数 n 是偶数，那么其展开式中间一项 $T_{\frac{n}{2}+1}$ 的二项式系数最大；如果 n 是奇数，那么其展开式中间两项 $T_{\frac{n+1}{2}}$ 与 $T_{\frac{n+1}{2}+1}$ 的二项式系数相等且最大．

此外，容易推出二项式系数满足的第 4 条性质．

4. 二项展开式的二项式系数的和等于 2^n．

在 $(1+x)^n=C_n^0x^n+C_n^1x^{n-1}+C_n^2x^{n-2}+\cdots+C_n^nx^0$ 中，令 $x=1$，则

$$\boxed{C_n^0+C_n^1+C_n^2+\cdots+C_n^n=2^n.}$$

例 1　证明在 $(a+b)^n$ 的展开式中，奇数项的二项式系数和等于偶数项的二项式系数和.

证明：在展开式

$$(a+b)^n=C_n^0a^n+C_n^1a^{n-1}b+C_n^2a^{n-2}b^2+C_n^3a^{n-3}b^3+\cdots+C_n^nb^n$$

中，令 $a=1$，$b=-1$，则得

$$(1-1)^n=C_n^0-C_n^1+C_n^2-C_n^3+\cdots+(-1)^nC_n^n,$$

即　$0=(C_n^0+C_n^2+C_n^4+\cdots)-(C_n^1+C_n^3+C_n^5+\cdots)$.

所以　$C_n^0+C_n^2+C_n^4+\cdots=C_n^1+C_n^3+C_n^5+\cdots$.

例 2　已知 $(x^2-1)^n$ 展开式的各项二项式系数和等于1 024，求展开式中含 x^6 的项.

解：因为展开式的各项二项式系数和等于 2^n，

所以　$2^n=1\ 024=2^{10}$，得到 $n=10$.

$(x^2-1)^{10}$展开式的通项为

$$T_{r+1}=C_{10}^r(x^2)^{10-r}\cdot(-1)^r=(-1)^rC_{10}^rx^{20-2r},$$

由已知，令 $20-2r=6$，求得 $r=7$，

所以所求项为

$$(-1)^7C_{10}^7x^6=-120x^6.$$

$(x^2-1)^n$ 展开式的各项系数和等于多少？

例 3　求$(1-x)^8$的展开式中二项式系数最大的项.

解：因为 $1-x$ 的幂指数 8 是偶数，由性质 3，$(1-x)^8$的展开式中间一项(即第5项)的二项式系数最大. 该项为

$$T_{\frac{8}{2}+1}=T_5=C_8^4(-x)^4=70x^4.$$

练习A

1. 填空：

(1) 已知 $C_{15}^5=a$，$C_{15}^9=b$，那么 $C_{16}^{10}=$______；

(2) 当 n 为偶数时，$(a+b)^n$ 展开式中，二项式系数最大项是第______项；
当 n 为奇数时，$(a+b)^n$ 展开式中，二项式系数最大项是第______项.

(3) 在$(2-\sqrt{x})^9$的展开式中，二项式系数最大项为______.

2. 求$(1-x)^{13}$的展开式中的含 x 的奇次项系数的和.

3. 证明：$C_n^0+C_n^2+C_n^4+\cdots+C_n^n=2^{n-1}$（$n$ 是偶数）.

4. 已知$\left(x^3+\dfrac{1}{x^2}\right)^n$的展开式中，只有第 6 项的系数最大，求展开式中的常数项.

练习B

填空：设$(3x-1)^8=a_8x^8+a_7x^7+\cdots+a_1x+a_0$，则

(1) $a_8+a_7+\cdots+a_1=$______；

(2) $a_8-a_7+a_6-a_5+a_4-a_3+a_2-a_1+a_0=$______；

(3) $a_8+a_6+a_4+a_2+a_0=$______.

习题1-3 A

1. 用杨辉三角展开$(a+b)^7$.

2. 用二项式定理展开：

(1) $(\sqrt{3}+\sqrt[3]{2})^6$；　　(2) $\left(\frac{\sqrt{x}}{2}-\frac{2}{\sqrt{x}}\right)^7$.

3. 化简：

(1) $(1+\sqrt{x})^5+(1-\sqrt{x})^5$；

(2) $(2x^{\frac{1}{2}}+3x^{-\frac{1}{2}})^4-(2x^{\frac{1}{2}}-3x^{-\frac{1}{2}})^4$.

4. (1) 求$(1-2x)^{15}$的展开式中前4项；

(2) 求$(2a^3-3b^2)^{10}$的展开式中第8项；

(3) 求$\left(1-\frac{1}{2x}\right)^{10}$的展开式中含$\frac{1}{x^5}$项的系数；

(4) 求$\left(2x^3+\frac{1}{2x^2}\right)^{10}$的展开式的常数项.

5. 求下列各指定项的二项式系数和系数：

(1) $(a-\sqrt[3]{b})^9$的展开式的第6项；

(2) $\left(\frac{2}{x^2}+x^3\right)^8$的展开式中含$x^4$的项.

6. (1) 已知$\left(\sqrt[3]{a^2}+\frac{1}{a}\right)^n$的展开式的第3项含有$a^2$，求$n$的值；

(2) 已知$(1+x)^n$的展开式中，第4项和第6项的系数相等，求这两项的系数.

7. 用二项式定理证明：

(1) $(n+1)^n-1$能被n^2整除；

(2) $99^{10}-1$能被100整除.

8. 求$C_{11}^1+C_{11}^3+C_{11}^5+\cdots+C_{11}^{11}$.

9. $(\sqrt[3]{x^{-1}}-\sqrt[5]{x^{-2}})^n$展开式的所有奇数项的系数和等于1 024，求展开式中二项式系

数最大项.

10. 求$(1+a)(1+b)^2(1+c)^3$展开式的各项系数和.

习题 1-3 B

1. (1) 已知$(\sqrt{x}+\sqrt[3]{x^2})^n$ 的展开式中，第 5 项的系数与第 3 项的系数之比是 7∶2，求展开式中含 x^5 的项；

(2) 已知$\left(\sqrt[4]{\dfrac{1}{a}}-\sqrt[3]{a^2}\right)^n$ 的展开式中，倒数第 3 项的系数的绝对值是 45，求展开式中含 a^3 的项.

2. 证明：

(1) $\left(x-\dfrac{1}{x}\right)^{2n}$的展开式的常数项是

$$(-2)^n\frac{1\cdot3\cdot5\cdot\cdots\cdot(2n-1)}{n!};$$

(2) $(1+2x)^{2n}$的展开式的中间一项是

$$\frac{1\cdot3\cdot5\cdot\cdots\cdot(2n-1)}{n!}(4x)^n.$$

3. 求$(1+x)^3+(1+x)^4+(1+x)^5+\cdots+(1+x)^{20}$展开式中含 x^3 项的系数.

4. 在$(\sqrt{2}+\sqrt[4]{3})^{50}$展开式中，有多少个有理项?

5. 填空：

(1) 多项式$(3x^4-x^3+2x-3)^8(3x-5)^4(7x^4-4x-2)^6$的展开式的各项系数和为______；

(2) 设 $(1-3x)^9=a_0+a_1x+a_2x^2+\cdots+a_9x^9$，则 $|a_0|+|a_1|+\cdots+|a_9|=$________.

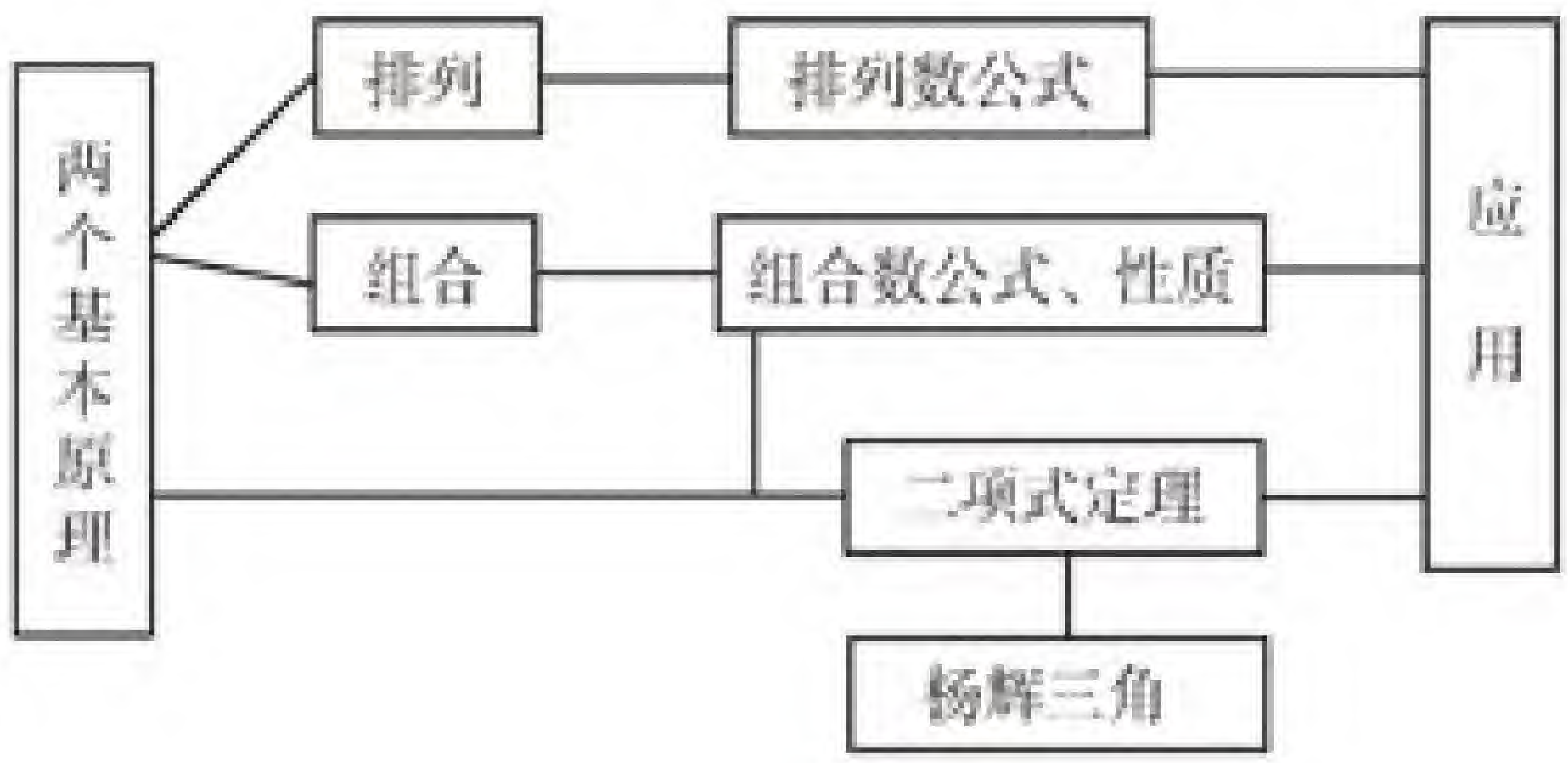
6. 求证：

(1) $C_n^1+2C_n^2+3C_n^3+\cdots+nC_n^n=n\cdot2^{n-1}$；(提示：$rC_n^r=nC_{n-1}^{r-1}$)

(2) $C_n^1-2C_n^2+3C_n^3-\cdots+(-1)^{n-1}nC_n^n=0$.

本章小结

Ⅰ 知识结构



Ⅱ 思考与交流

1. 分类加法计数原理和分步乘法计数原理的根本区别是什么？在应用中各应该注意什么问题？学习完排列组合后，你是否已经体会到了两个基本原理所起的重要作用？请谈谈你的体会.

2. 排列问题与组合问题有何不同？它们之间的相互关系如何？排列与组合的不同点以及相互之间的联系，在公式 $A_n^m=C_n^m\cdot A_m^m$ 中是怎样体现的？

3. 请你小结排列组合应用题的类型和解题方法.

4. $(a+b)^n$ 展开式中的系数为什么能以组合数的形式表示？二项式系数 $C_n^r=f(r)(0\leqslant r\leqslant n，r\in\mathbf{N})$是 r 的函数，请以函数的语言叙述 C_n^r 的性质.

5. 在应用二项展开式的通项公式解题时，需要注意什么问题？常犯的错误有哪些？

Ⅲ 巩固与提高

1. 填空：

（1）乘积$(a_1+a_2+\cdots+a_m)(b_1+b_2+\cdots+b_n)$展开后，共有______项；

(2) 如果$\dfrac{C_n^m}{C_{n'}^{m}}=a$，则$\dfrac{A_n^m}{A_{n'}^{m}}=$______；

(3) 已知 $C_{10}^x=C_{10}^3$，则 $x=$______；

(4) 安排 6 名歌手的演出顺序时，要求某歌手既不第一个出场，也不最后一个出场，共有______种不同的排法；

(5) 学生从本学期开设的 5 门选修课中任选 3 门，从 4 种课外活动小组中任选 1 个，则该学生共有______种不同的选择方法.

2. 求证：

(1) $\dfrac{(2n)!}{2^n\cdot n!}=1\cdot 3\cdot 5\cdot\cdots\cdot(2n-1)$；

(2) $1\cdot 1!+2\cdot 2!+3\cdot 3!+\cdots+n\cdot n!=(n+1)!-1$.

3. (1) 已知$\dfrac{1}{C_5^x}-\dfrac{1}{C_6^x}=\dfrac{7}{10C_7^x}$，求 C_8^x；

(2) 已知$\dfrac{C_n^{m-1}}{2}=\dfrac{C_n^m}{3}=\dfrac{C_n^{m+1}}{4}$，求 n 与 m.

4. (1) 一个集合由 8 个元素组成，这个集合含有 3 个元素的子集有多少个？

(2) 一个集合由 5 个元素组成，这个集合至多含有 3 个元素的子集有多少个？

5. 由数字 1，2，3，4，5，6 可组成多少个没有重复数字的：

(1) 正整数？

(2) 比 50 000 大的正整数？

(3) 五位奇数？

6. 有 6 名同学站成一排，符合下列各题要求的不同排法共有多少种？

(1) 甲同学不站在排头；

(2) 甲、乙、丙三位同学两两不相邻；

(3) 甲、乙两同学不相邻，且乙、丙两同学也不相邻；

(4) 甲、乙两同学相邻，且丙、丁两同学也相邻.

7. (1) 平面内有 n 条直线，其中没有两条直线相互平行，也没有三条直线相交于一点，一共有多少个交点？

(2) 空间有 n 个平面，其中没有两个平面相互平行，也没有三个平面相交于一条直线，一共有多少条交线？

8. 100 件产品中有 97 件合格品，3 件次品，从中任意抽取 5 件进行检查：

(1) 都是合格品的抽法有多少种？

(2) 恰好有 2 件次品的抽法有多少种？

(3) 至少有 2 件是次品的抽法有多少种？

9. 书架上有 4 本不同的数学书，5 本不同的物理书，3 本不同的化学书，将其全部竖起排成一排：

(1) 如果不使同类的书分开，一共有多少种不同的排法？

(2) 如果使物理书两两不相邻，一共有多少种不同的排法？

10. (1) 将 6 名应届大学毕业生分给 2 个用人单位，每个单位至少 2 名，一共有多少种不同的分配方案？

(2) 某公司将 6 个招聘名额分给 3 个下属单位，一个单位 3 个名额，一个单位 2 个名额，一个单位 1 个名额，一共有多少种不同的分配方案？

11. 某电脑用户计划用不超过 500 元的资金购买单价分别为 60 元、70 元的单片软件和盒装磁盘，根据需要软件至少买 3 片，磁盘至少买 2 盒，一共有多少种不同的购买方式？

12. (1) 求$(a+\sqrt{b})^{12}$展开式中的第 9 项；

(2) 求$(1-2x)^5(1+3x)^4$展开式中，按 x 的升幂排列的前三项；

(3) 求$\left(9x-\dfrac{1}{3\sqrt{x}}\right)^{18}$展开式的常数项，并说明它是展开式的第几项；

(4) 求$(1+x+x^2)(1-x)^{10}$展开式中 x^4 的系数；

(5) 求$\left(|x|+\dfrac{1}{|x|}-2\right)^3$展开式的常数项.

13. 已知$(1+\sqrt{x})^n$ 的展开式中第 9 项、第 10 项、第 11 项的二项式系数成等差数列，求含 x^2 的项.

14. (1) 用二项式定理证明 $55^{55}+9$ 能被 8 整除；

(2) 用二项式定理求 89^{10}除以 88 的余数.

15. 证明：$(1+x)^{2n}$展开式中 x^n 的系数等于$(1+x)^{2n-1}$展开式中 x^n 的系数的 2 倍.

16. 已知$\left(\sqrt{x}+\dfrac{1}{\sqrt[3]{x}}\right)^n$ 展开式的二项式系数之和比$(a+b)^{2n}$展开式的二项式系数之和小 240，求：

(1) $\left(\sqrt{x}+\dfrac{1}{\sqrt[3]{x}}\right)^n$ 展开式的第 3 项；

(2) $(a+b)^{2n}$展开式中的系数最大项.

17. 求证：$(C_n^0)^2+(C_n^1)^2+(C_n^2)^2+\cdots+(C_n^n)^2=\dfrac{(2n)!}{n!\cdot n!}$.

(提示：利用$(1+x)^n(1+x)^n=(1+x)^{2n}$中，左右两边含 x^n 项的系数相等.)

18. 求$(a^{\frac{1}{3}}b^{-\frac{1}{2}}+a^{-\frac{1}{2}}b^{\frac{1}{3}})^{11}$展开式中 a 和 b 的指数相等的项.

IV 自测与评估

1. 选择题：

(1) 5 个人分 4 本同样的书，每人至多一本，而且必须分完，那么不同分法的种数是(　　).

(A) 5^4　　　　(B) 4^5

(C) $5\times4\times3\times2$　　(D) $\dfrac{5\times4\times3\times2}{4!}$

(2) 一个文娱团体下基层进行宣传演出，准备的节目表中原有 4 名歌手演唱，如果保持着 4 名歌手演唱的相对顺序不变，拟再添加 2 个小品节目，则不同的节目表可排出（　　）.

(A) 20 种　　(B) 25 种

(C) 30 种　　(D) 32 种

(3) 5 个身高均不相同的学生排成一排合影留念，最高个子站在中间，从中间到左边和从中间到右边一个比一个矮，则这样的排法共有（　　）.

(A) 6 种　　(B) 8 种

(C) 12 种　　(D) 16 种

(4) 若$(1+2x)^6$的展开式中第 2 项大于它的相邻两项，则 x 的取值范围是（　　）.

(A) $\left(\dfrac{1}{12},\ \dfrac{1}{5}\right)$　　(B) $\left(\dfrac{1}{6},\ \dfrac{1}{5}\right)$

(C) $\left(\dfrac{1}{12},\ \dfrac{2}{3}\right)$　　(D) $\left(\dfrac{1}{6},\ \dfrac{2}{5}\right)$

(5) 在$\left(\sqrt[3]{2}x-\dfrac{1}{\sqrt{2}}\right)^{20}$的展开式中系数是正有理数的项共有（　　）.

(A) 4 项　　(B) 5 项

(C) 6 项　　(D) 7 项

2. 填空题：

(1) 3 名医生和 6 名护士被分配到 3 所学校为学生体检，每校分配 1 名医生和 2 名护士，共有______种不同的分配方案；

(2) 4 个人站成一列，重新站队时各人都不站在原来的位置上，共有______种不同的站法；

(3) 在$(ax+1)^7$展开式中，x^3 的系数是 x^2 的系数与 x^4 的系数的等差中项，若 $a>1$，那么 $a=$______.

3. 某小组有 3 名女生、4 名男生，从中选出 3 名代表，要求至少女生与男生各有一名，共有多少种不同的选法？(要求用两种方法求解)

4. 现有分别印有 0，1，3，5，7，9 六个数字的六张卡片，如果允许 9 可以当 6 使用，那么从中任意抽出三张，可以组成多少个不同的三位数？

5. 一个口袋里装有 4 个不同的红球，6 个不同的白球．若取出一个红球记两分，取出一个白球记一分，从口袋中取出 5 个球，使总分不低于七分的取法共有多少种？

6. 已知$\left(\sqrt{x}+\dfrac{1}{x}\right)^n$ 的展开式中，第 4 项的二项式系数与第 5 项的二项式系数之比为 1∶3，求二项式系数最大的项.

数　学　文　化

阅读与欣赏

$x_1+x_2+\cdots+x_r=n$ 非负整数解的个数

排列、组合问题和下面这个代数问题是有联系的：求满足方程

$$x_1+x_2+\cdots+x_r=n$$

的非负整数解$(x_1,\ x_2,\ \cdots,\ x_r)$的个数.

考虑由 n 个 1 与$(r-1)$个 0 组成的排列. 将每一个这样的排列与方程的一个解$(x_1,\ x_2,\ \cdots,\ x_r)$按如下方式对应起来：使 x_1 等于排列中第一个 0 左边 1 的个数，x_2 等于第一个 0 与第二个 0 之间 1 的个数，x_3 等于第二个 0 与第三个 0 之间 1 的个数，如此继续直到 x_r，它等于最后一个 0 右边 1 的个数. 例如，若 $n=6$，$r=4$，则排列(1，1，0，0，1，1，1，0，1)对应着解 $x_1=2$，$x_2=0$，$x_3=3$，$x_4=1$. 显然在 n 个 1 与$(r-1)$个 0 组成的所有排列与方程的全体解之间的这种对应是一一对应. 由于 n 个 1 与$(r-1)$个 0 组成的排列共有 C_{n+r-1}^{n} 个，所以方程共有 C_{n+r-1}^{n} 个非负整数解$(x_1,\ x_2,\ \cdots,\ x_r)$.

把 8 个相同的篮球任意分发给甲、乙、丙、丁 4 所中学，试问不同的分法共有多少种？

设甲、乙、丙、丁 4 所中学分到的篮球数分别为 x_1，x_2，x_3，x_4. 问题就是求方程

$$x_1+x_2+x_3+x_4=8$$

的非负整数解的个数.

由上所述，这些非负整数解的个数是

$$C_{8+4-1}^{8}=C_{11}^{8}=C_{11}^{3}=\frac{11\times10\times9}{3\times2\times1}=165,$$

它也就是问题所求的分法种数.

进一步问，每所中学至少分到 1 个篮球的不同分法有多少种？

我们可以先把 4 个篮球平分给这 4 所中学，然后把余下的 4 个篮球任意分发，类似于上面的讨论，容易得到每所中学至少分到 1 个篮球的分法种数是

$$C_{4+4-1}^{4}=C_{7}^{4}=C_{7}^{3}=\frac{7\times6\times5}{3\times2\times1}=35.$$

如果同学们去直接罗列 8 个篮球分给 4 所中学的各种情况，很快就会发现，这绝不是一件轻松的事情.

第二章　概　率

2.1 离散型随机变量及其分布列

2.2 条件概率与事件的独立性

2.3 随机变量的数字特征

2.4 正态分布

我们已学过概率的初步知识，对现实世界中随机现象的不确定性有了初步认识，但要描述随机现象在大量试验中的统计规律性，还需要进一步学习一些概率知识．让我们先看几个实际问题．

快讯：雅典时间 8 月 14 日 11 点，中国选手杜丽在雅典奥运会女子 10 米气步枪决赛中技压群雄，总成绩 502 环，打破奥运会记录．夺得了第 28 届雅典奥运会的首枚金牌．同时这也是中国代表团在本届奥运会上夺得的首枚金牌，中国奥运代表团取得开门红……（新华网 8 月 14 日电）

读完快讯，我们也许会问：射击运动员的射击水平应如何描述？

新华网华盛顿 4 月 6 日电（记者单磊）美国哥伦比亚广播电台网站根据自己的统计，将本赛季 NBA 的新秀进行排名……姚明本赛季为休斯顿火箭队打了 76 场比赛，平均每场比赛上场 29.2 分钟，得分 13.8 分，篮板球 8.2 个，助攻 1.7 次，投篮命中率 50.7%，罚篮命中率 80.9%．姚明的综合得分是 74.9 分．

看了报道，你可能会想到一些技术性问题：如果姚明在某节比赛中得到 4 次罚球机会，并假定每次投篮都互不影响，那么他恰好投中 3 次的可能性有多大？4 罚投中 3 次以上的可能性又有多大？

甲、乙、丙三个同学得到一张电影票，为了公平起见，三个人通过抓阄决定谁去看电影．甲先抓，乙其次，剩下的阄归丙，丙同学提出疑问，这样抓阄对自己公平吗？

上面提出的一些问题你能回答吗？试着给出你的答案．

本章将从实例中引入随机变量及其分布的概念，超几何分布的实际应用．在讲述条件概率与事件独立性的基础上，介绍二项分布及其应用．研究离散型随机变量的两个数字特征：期望与方差．最后简单介绍概率论中极其重要的分布——正态分布．

2.1 离散型随机变量及其分布列

2.1.1 离散型随机变量

选手每次射击时，可能击中靶心，也可能击中靶心周围的区域，所以这是一个随机现象．在射击比赛中，选手击中靶上的圆形或环形区域内得分，得分值由靶心往外依次可记为：10 环，9 环，8 环，…，1 环．选手击中靶上最大圆以外区域或脱靶的分值都记为 0 环．

我们以前研究过的许多随机现象的每一个可能的结果，如抛掷骰子所得到的点数，体育彩票开奖得出的号码，气象台对明天天气最高温度的预报，都是一些数量，都可以用一个变量来表示．

有些随机现象的结果虽然不是数量，但可以将它数量化．例如抛一枚硬币，所有可能的结果是：

$$A_1=\text{“正面向上”与}A_2=\text{“反面向上”}.$$

为了在数学上描述这些可能的结果，可以用“1”代表正面向上，用“0”代表反面向上．

在这些试验中，试验可能出现的结果可以用一个变量 X 来表示，并且 X 是随着试验的结果的不同而变化的，我们把这样的变量 X 叫做一个**随机变量**．随机变量常用大写字母 X，Y，…表示．[1]

注

[1] 也可用希腊字母 ξ，η，…表示．

例如，设某射击选手每次射击所得的环数是 X，那么 X 是一个随机变量．X 的取值范围是

$$\{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10\}.$$

则 $X=0$，表示射中 0 环；$X=1$，表示射中 1 环……$X=10$，表示射中 10 环．

又如，100 件产品中，含有 5 件次品，从中取出 4 件，那么可能出现的“次品件数” X 是一个随机变量，X 的取值范围是 $\{0,\ 1,\ 2,\ 3,\ 4\}$．

如果随机变量 X 的所有可能的取值都能一一列举出来，则称 X 为**离散型随机变量**．上面所举出的都是离散型随机变量的例子．我们本章也主要来学习离散型随机变量．

练习A

1．写出下列各离散型随机变量可能取的值：

（1）从 10 张已编号的卡片（1～10号）中任取一张，被取出的卡片的号数；

（2）抛掷一个骰子得到的点数；

（3）一个袋子里装有 5 个白球和 5 个黑球．从中任取 3 个，其中所含白球的个数；

（4）同时抛掷 5 枚硬币，得到硬币反面向上的个数．

2. 把一枚硬币先后抛掷两次．如果出现两个正面得 5 分，出现两个反面得 -3 分，其他结果得 0 分．用 X 来表示得到的分值，列表写出可能出现的结果与对应的 X 值．

练习B

假设进行一次从袋中摸出一个球的游戏，袋中有 3 个红球、4 个白球、1 个蓝球、2 个黑球，摸到红球得 2 分、白球得 0 分、蓝球得 1 分、黑球得 -2 分，试列表写出可能的结果、对应的分值 X 及相应的概率．

2.1.2 离散型随机变量的分布列

对于一个离散型随机变量来说，我们不仅要知道它可能取哪些值，更重要的是要知道它取各个值的概率分别有多大，这样才能对这个离散型随机变量有较深入的了解．例如，在射击问题里，我们只有知道命中环数为 0，1，2，…，10 的概率分别是多少，才能了解选手的射击水平有多高．根据某个选手在一段时间里的成绩，可以得到下表：

命中环数 X	0	1	2	3	4	5	6	7	8	9	10
概率 P	0	0	0.01	0.01	0.02	0.02	0.06	0.09	0.28	0.29	0.22

上面的表格，使我们对选手的射击水平有了一个比较全面的了解．这个例子表明，要掌握一个离散型随机变量 X 的取值规律，必须知道：

（1）X 所有可能取的值 x_1，x_2，…，x_n；

（2）X 取每一个值 x_i 的概率 p_1，p_2，…，p_n．

这就是说，需要列出下表：

X	x_1	x_2	…	x_i	…	x_n
P	p_1	p_2	…	p_i	…	p_n

我们称这个表为离散型随机变量 X 的**概率分布**，或称为离散型随机变量 X 的**分布列**．由

分布列能一目了然地看出随机变量 X 的取值范围及取这些值的概率.

例如，通过上面射击选手的命中环数 X 的分布列，可以全面了解这名选手的射击成绩的概率分布情况. 这名选手命中 10 环的概率为 $P(X=10)=0.22$；没有命中 10 环的概率是多少呢？它是命中 10 环的对立事件，因此，$P(X\neq 10)=1-P(X=10)=0.78$；事件"命中 9 环"和事件"命中 10 环"不可能同时发生，为互斥事件，所以命中的环数大于 8 环的概率为 $P[(X=9)\cup(X=10)]=P(X=9)+P(X=10)=0.29+0.22=0.51$.

计算一下上面表中选手命中环数对应的概率值的和，不难发现各 p_i 值的和等于 1. 在一般情况下，因为基本事件空间

$$\Omega=(X=x_1)\cup(X=x_2)\cup\cdots\cup(X=x_i)\cup\cdots\cup(X=x_n)$$

是一个必然事件，上面右式各项彼此互斥，根据互斥事件的概率加法公式有

$$\begin{aligned}1=P(\Omega)&=P[(X=x_1)\cup(X=x_2)\cup\cdots\cup(X=x_i)\cup\cdots\cup(X=x_n)]\\&=P(X=x_1)+P(X=x_2)+\cdots+P(X=x_i)+\cdots+P(X=x_n)\\&=p_1+p_2+\cdots+p_i+\cdots+p_n,\end{aligned}$$

所以离散型随机变量的分布列有下面两条性质：

(1) $p_i\geqslant 0$，$i=1，2，3，\cdots，n$；

(2) $p_1+p_2+\cdots+p_n=1$.

例 1 篮球运动员在比赛中每次罚球命中得 1 分，不中得 0 分. 已知某运动员罚球命中的概率为 0.7，求他罚球一次的得分的分布列.

解：用随机变量 X 表示"每次罚球得的分值"，根据题意，X 可能的取值为 0，1，且取这两个值的概率分别为 0.7，0.3，因此所求的分布列是

X	1	0
P	0.7	0.3

如果随机变量 X 的分布列为

X	1	0
P	p	q

其中 $0<p<1$，$q=1-p$，则称离散型随机变量 X 服从参数为 p 的**二点分布**. 例 1 中篮球运动员每次罚球得的分值服从 $p=0.7$ 的二点分布.

例 2 掷一颗骰子，所掷出的点数为随机变量 X：

(1) 求 X 的分布列；

(2) 求"点数大于 4"的概率；

(3) 求"点数不超过 5"的概率.

解：(1) X 的分布列为

> 为了简便，例 2 中的事件"点数大于 4"可以表示为"$X>4$"，是指互斥事件"$X=5$""$X=6$"的和，根据互斥事件的概率加法公式，可以求出掷一颗骰子，"所掷出的点数大于 4"的概率.

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(2) $P(X>4)=P(X=5)+P(X=6)=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$；

(3) $P(X\leqslant 5)=P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)=5\cdot\frac{1}{6}=\frac{5}{6}$.

例 3 某同学向图 2-1 所示的圆形靶投掷飞镖，飞镖落在靶外的概率为 0.1，飞镖落在靶内的各个点是随机的．已知圆形靶中三个圆为同心圆，半径分别为 30 cm，20 cm，10 cm，飞镖落在不同区域的环数如图中标示．设这位同学投掷一次得到的环数这个随机变量为 X，求 X 的分布列．

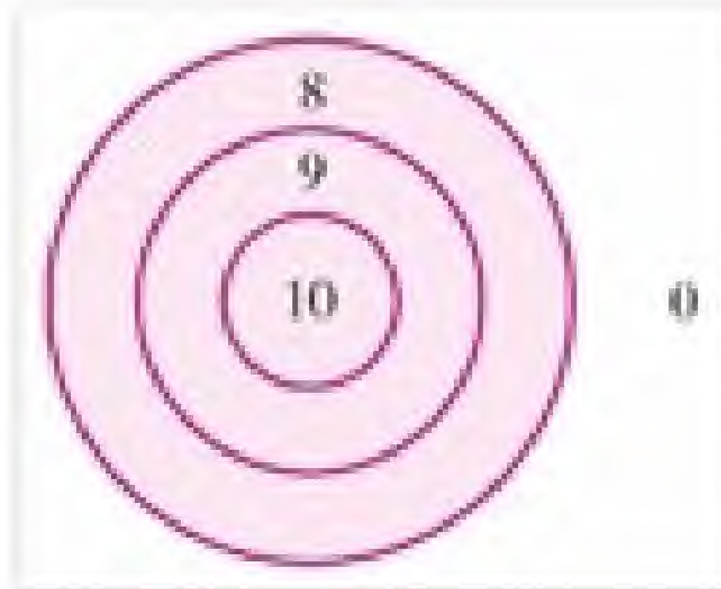


图 2-1

解：由题意可知，飞镖落在靶内各个区域的概率与它们的面积成正比，而与它们的位置和形状无关．由圆的半径值可得到三个同心圆的半径比为 3∶2∶1，面积比为9∶4∶1，所以 8 环区域，9 环区域，10 环区域的面积比为 5∶3∶1，则掷得 8 环，9 环，10 环的概率可分别设为 $5k$，$3k$，k，根据离散型随机变量分布列的性质(2)有

$$0.1+5k+3k+k=1,$$

解得　$k=0.1$．得到离散型随机变量 X 的分布列为

X	0	8	9	10
P	0.1	0.5	0.3	0.1

练习A

1. 下面列出的表格是否是某个离散型随机变量的分布列？试用分布列的性质加以说明．

(1)

X	−1	0	1	2	3
P	0.2	0.2	0.2	0.2	0.2

(2)

X	0	1	2	3	4	5
P	0.1	−0.2	0.3	0.4	0.2	0.2

2. 投掷一枚硬币，设 $X=\begin{cases}1, & \text{出现正面} \\ 0, & \text{出现反面}\end{cases}$ 求随机变量 X 的分布列.

3. 一个布袋中共有 50 个完全相同的球，其中标记为 0 号的 5 个，标记为 n 号的分别有 n 个（$n=1, 2, \cdots, 9$）. 求从袋中任取一球所得球号数的分布列.

4. 掷两颗骰子，设掷得的点数和为随机变量 X：

   (1) 求 X 的分布列；

   (2) 求“点数和大于 9”的概率；

   (3) 求“点数和不超过 7”的概率.

练习B

1. 在 8 张扑克牌中，有“黑桃，红心，梅花，方块”这四种花色的牌各两张. 从中任取两张，求其中取得黑桃花色牌的张数的分布列.

2. 某商店购进一批西瓜，预计晴天西瓜畅销，可获利 1 000 元；阴天销路一般，可获利 500 元；下雨天西瓜滞销，这时将亏损 500 元. 根据天气预报，未来数日晴天的概率为 0.4，阴天的概率为 0.2，下雨的概率为 0.4，试写出销售这批西瓜获利的分布列.

2.1.3 超几何分布

某校组织一次认识大自然夏令营活动，有 10 名同学参加，其中有 6 名男生、4 名女生，为了活动的需要，要从这 10 名同学中随机抽取 3 名同学去采集自然标本，那么其中恰有 1 名女生的概率有大?

从 10 名同学中随机抽取 3 名同学，考察所有可能的结果，这是一个随机试验. 由于是随机抽取，任意 3 名同学被选中的可能性相等，所以这个试验是一个古典概型问题，基本事件空间是所有可能的抽取结果. 根据组合数知识，从 10 名同学中任意选出 3 名同学，共有 $C_{10}^3=120$ 种不同的选法，所以基本事件空间包含的基本事件总数为 120 个. 其中恰好有 1 名女生就意味着选出 1 名女生和 2 名男生. 由分步乘法计数原理得到恰有 1 名女生的基本事件为 $C_4^1C_6^2=60$ 个，因此其中恰有 1 名女生的概率为

$$P(\text{“恰有 1 名女生”})=\frac{C_4^1C_6^2}{C_{10}^3}=\frac{60}{120}=\frac{1}{2}.$$

采集标本的同学都是女生的概率有多大呢？类似可得到恰有3名女生的基本事件个数为$C_4^3C_6^0=4$个，因此恰有3名女生的概率为

$$P(\text{“恰有 3 名女生”})=\frac{C_4^3C_6^0}{C_{10}^3}=\frac{4}{120}=\frac{1}{30},$$

可见结果全部是女生的概率要小得多．其他抽取结果概率的计算也与上面两式类似．

实际生活中有很多像上面例子这样的问题．一般地，设有总数为N件的两类物品，其中一类有M件，从所有物品中任取n件（$n\leqslant N$），这n件中所含这类物品件数X是一个离散型随机变量，它取值为m时的概率为

$$P(X=m)=\frac{C_M^mC_{N-M}^{n-m}}{C_N^n}\ (0\leqslant m\leqslant l,\ l\text{ 为 }n\text{ 和 }M\text{ 中较小的一个}).\qquad ①$$

我们称离散型随机变量X的这种形式的概率分布为**超几何分布**，也称X服从参数为N，M，n的超几何分布．

在超几何分布中，只要知道N，M和n，就可以根据公式①求出X取不同m值时的概率$P(X=m)$，从而列出X的分布列．

例1 在一个口袋中装30个球，其中有10个红球，其余为白球，这些球除颜色外完全相同．游戏者一次从中摸出5个球，摸到且只能摸到4个红球就中一等奖．那么获一等奖的概率有多大（结果保留两位有效数字）？

解：根据题意，设随机变量X表示摸出红球的个数，则X服从参数为$N=30$，$M=10$，$n=5$的超几何分布．X可能的取值为0，1，2，3，4，5．由题目可知，要求摸到4个红球的概率，根据公式①可得摸到4个红球的概率为

$$P(X=4)=\frac{C_{10}^4C_{30-10}^{5-4}}{C_{30}^5}=\frac{4\ 200}{142\ 506}\approx 0.029.$$

因此获一等奖的概率约为0.029.

例2 一批产品共100件，其中有5件次品．现在从中任取10件检查，求取到的次品件数的分布列（精确到0.000 01）.

解：根据题意，取到的次品件数X为离散型随机变量，且X服从参数为$N=100$，$M=5$，$n=10$的超几何分布．它可能的取值为0，1，2，3，4，5，根据公式①算出其相应的概率依次为

$$P(X=0)=\frac{C_5^0C_{100-5}^{10}}{C_{100}^{10}}\approx 0.583\ 75,$$

$$P(X=1)=\frac{C_5^1C_{100-5}^{10-1}}{C_{100}^{10}}\approx 0.339\ 39,$$

$$P(X=2)=\frac{C_5^2C_{100-5}^{10-2}}{C_{100}^{10}}\approx 0.070\ 22,$$

$$P(X=3)=\frac{C_5^3C_{100-5}^{10-3}}{C_{100}^{10}}\approx 0.006\ 38,$$

$$P(X=4)=\frac{C_5^4C_{100-5}^{10-4}}{C_{100}^{10}}\approx 0.000\ 25,$$

$$P(X=5)=1-P(X=0)-P(X=1)-P(X=2)-P(X=3)-P(X=4)\approx 0.000\ 01.$$

因此 X 的分布列为

X	0	1	2	3	4	5
P	0.583 75	0.339 39	0.070 22	0.006 38	0.000 25	0.000 01

练习A

1. 一批产品共 100 件，次品率为 4%，从中任意抽取 10 件检查，求抽得的次品数的分布列.

2. 盒中有 16 个白球和 4 个黑球，从中任意取出 3 个，设 X 表示其中黑球的个数，求出 X 的分布列.

练习B

若 M 件产品中包含 m（$m\geqslant3$）件次品，$M-m\geqslant3$，从中任意取出 3 件，设 X 表示取出的次品数，列出 X 的分布列.

习题 2-1 A

1. 写出下列各离散型随机变量可能取的值：

（1）从 7 张已编号的卡片（1～7 号）中任取两张，被取出的卡片的号数之和；

（2）从含有 5 件次品的一批产品中任取一件，被取到的次品数；

（3）一个袋子里装有 4 个白球、5 个黑球和 6 个黄球. 从中任取 4 个，其中所含黑球的个数；

（4）先后抛掷红、蓝两个骰子得到的点数之和.

2. 从 1，2，3，4 四个数字中，任意地取出两数，求取出的两数之和的分布列.

3. 从含有 4 件次品的 50 件产品中任取 5 件，求抽得的次品数的分布列.

4. 在 10 个乒乓球中有 8 个正品，2 个次品. 从中任取 3 个，求其中所含次品数的分布列.

习题 2-1 B

1. 箱子中装有 50 个苹果，其中有 40 个是合格品，10 个是次品. 从箱子中任意抽取 10 个苹果，其中的次品数为随机变量 X，列出 X 的分布列，并求出

（1）所抽取的 10 个苹果中没有次品的概率；

（2）所抽取的 10 个苹果中恰有 2 个次品的概率；

（3）所抽取的 10 个苹果中最多有 2 个次品的概率.

2. 5 张卡片上分别标有号码 1，2，3，4，5. 从中任取 3 张，求 3 张卡片中最大号码的分布列.

2.2 条件概率与事件的独立性

2.2.1 条件概率

在很多实际问题中，需要考虑一个事件在“某事件已发生”这个附加条件下的概率. 我们来看下面的问题.

抛掷红、蓝两颗骰子. 设

事件 $A=$ “蓝色骰子的点数为 3 或 6”，

事件 $B=$ “两颗骰子的点数之和大于 8”.

我们用 x 代表抛掷红骰子所得到的点数，用 y 代表抛掷蓝骰子所得到的点数，则这个试验的基本事件空间为$S=\{(x, y)\mid x\in \mathbf{N}, y\in \mathbf{N}, 1\leqslant x\leqslant 6, 1\leqslant y\leqslant 6\}$. 作图 2-2，容易看出，基本事件空间的元素与图中的点一一对应. 所以抛掷红、蓝两颗骰子这一试验的基本事件总数为 36. 事件 B 所包含的基本事件对应图中三角实线所包围的点，个数为 10. 所以，事件 B 发生的概率 $P(B)=\dfrac{10}{36}$.

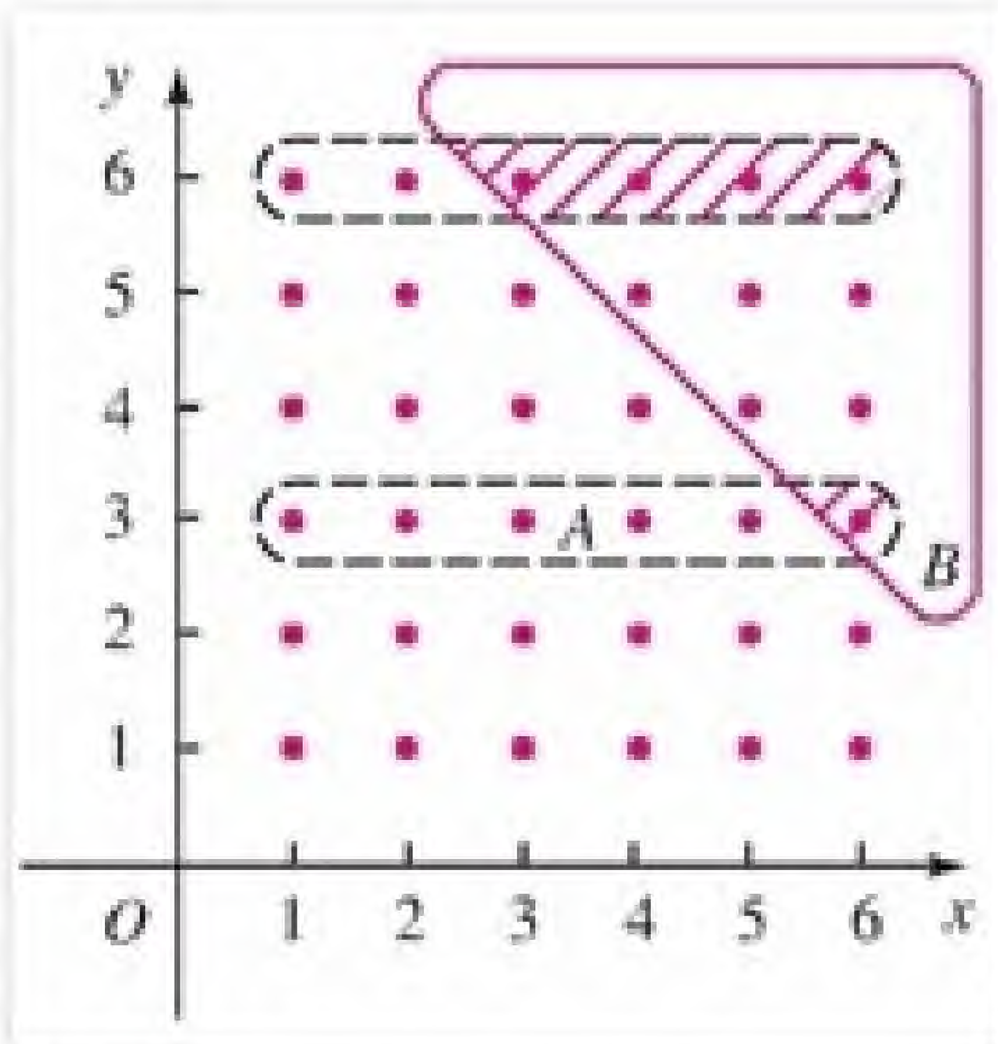


图 2-2

当已知蓝色骰子的点数为 3 或 6 时，事件 B 发生的概率是多少呢？也就是说，要求事件 B 在“事件 A 已发生”这个附加条件下的概率是多少. 事件 A 已发生的所有可能的结果对应图中长条虚线所包围的 12 个点，其中阴影部分的 5 个点的“点数之和大于 8”. 所以事件 B 在“事件 A 已发生”条件下的概率是$\dfrac{5}{12}$.

从这个例子中看到，事件 B 在“事件 A 已发生”这个附加条件下的概率与没有这个附加条件的概率是不同的. 对于任何两个事件 A 和 B，在已知事件 A 发生的条件下，事件 B 发生的概率叫做**条件概率**，用符号“$P(B|A)$”来表示.

在图 2-2 中阴影部分的 5 个点对应的事件为“事件 A 发生并且事件 B 也发生”，我们把由事件 A 和 B 同时发生所构成的事件 D，称为事件 A 与 B 的**交**（或**积**），记做 $D=A\cap B$（或 $D=AB$）.

容易得到上面的例子中 $P(A)=\dfrac{12}{36}$，$P(A\cap B)=\dfrac{5}{36}$，而

$$P(B|A)=\frac{5}{12}=\frac{\frac{5}{36}}{\frac{12}{36}}=\frac{P(A\cap B)}{P(A)}.$$

一般地，我们有条件概率公式

$$P(B|A)=\frac{P(A\cap B)}{P(A)},\ P(A)>0. \qquad ②$$

例 1 一个家庭中有两个小孩. 假定生男、生女是等可能的. 已知这个家庭有一个是女孩，问这时另一个小孩是男孩的概率是多少?

解：一个家庭的两个小孩子只有 4 种可能：{两个都是男孩}，{第一个是男孩，第二个是女孩}，{第一个是女孩，第二个是男孩}，{两个都是女孩}，由题目假定可知这 4 个基本事件发生是等可能的. 根据题意，设基本事件空间为 Ω，$A=$ "其中一个是女孩"，$B=$ "其中一个是男孩". 则

$\Omega=\{$(男，男)，(男，女)，(女，男)，(女，女)$\}$，

$A=\{$(男，女)，(女，男)，(女，女)$\}$，

$B=\{$(男，男)，(男，女)，(女，男)$\}$，

$A\cap B=\{$(男，女)，(女，男)$\}$.

问题是求在事件 A 发生的情况下，事件 B 发生的概率，即求 $P(B|A)$. 由上面分析可知 $P(A)=\frac{3}{4}$，$P(A\cap B)=\frac{2}{4}$.

由公式②可得

$$P(B|A)=\frac{\frac{2}{4}}{\frac{3}{4}}=\frac{2}{3}.$$

因此所求条件概率为$\frac{2}{3}$.

例 2 设某种动物由出生算起活到 20 岁的概率为 0.8，活到 25 岁的概率为 0.4，现有一个 20 岁的这种动物，问它能活到 25 岁的概率是多少?

解：设 $A=$ "能活到 20 岁"，$B=$ "能活到 25 岁"，则 $P(A)=0.8$，$P(B)=0.4$，而所求概率为 $P(B|A)$，由于$B\subseteq A$，故 $A\cap B=B$，于是

$$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{P(B)}{P(A)}=\frac{0.4}{0.8}=0.5,$$

所以这个动物能活到 25 岁的概率是 0.5.

例 3 甲、乙两地都位于长江下游，根据一百多年的气象记录，知道甲、乙两地一年中雨天占的比例分别为 20%和 18%，两地同时下雨的比例为 12%，问：

(1) 乙地为雨天时甲地也为雨天的概率是多少?

(2) 甲地为雨天时乙地也为雨天的概率是多少?

解：设 $A=$ "甲地为雨天"，$B=$ "乙地为雨天"，则根据题意有

$$P(A)=0.20,\ P(B)=0.18,\ P(A\cap B)=0.12,$$

所以

(1) 乙地为雨天时甲地也为雨天的概率是

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{0.12}{0.18}=0.67.$$

（2）甲地为雨天时乙地也为雨天的概率是

$$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{0.12}{0.20}=0.60.$$

练习A

1. 把一枚硬币任意抛掷两次，事件 A= “第一次出现正面”，事件 B= “第二次出现正面”，求 $P(B|A)$.
2. 抛掷红、蓝两个骰子，事件 A= “红骰子出现 4 点”，事件 B= “蓝骰子出现的点数是偶数”，求 $P(A|B)$.
3. 盒子中有 25 个外形相同的球，其中 10 个白的，5 个黄的，10 个黑的，从盒子中任意取出一球，已知它不是黑球，试求它是黄球的概率.
4. 设某种灯管使用了 500 h 还能继续使用的概率是 0.94，使用到 700 h 后还能继续使用的概率是 0.87，问已经使用了 500 h 的灯管还能继续使用到 700 h 的概率是多少？

练习B

1. 假定生男孩或生女孩是等可能的．在一个有 3 个孩子的家庭中，已知有一个男孩，求至少有一个女孩的概率.
2. 掷两枚均匀的骰子，已知点数不同，求至少有一个是 6 点的概率.

2.2.2 事件的独立性

我们知道，当事件 A 的发生对事件 B 的发生有影响时，条件概率 $P(B|A)$和概率 $P(B)$一般是不相等的，但有时事件 A 的发生看上去对事件 B 的发生没有影响，比如依次抛掷两枚硬币，抛掷第一枚硬币的结果(事件 A)应该对第二枚硬币的结果(事件 B)没有影响，这时 $P(B|A)$与$P(B)$相等吗？

让我们先来看一个例子.

例 1 在大小均匀的 5 个鸡蛋中有 3 个红皮蛋，2 个白皮蛋，每次取一个，有放回地

取两次，求在已知第一次取到红皮蛋的条件下，第二次取到红皮蛋的概率.

解：设 $A=$ “第一次取到红皮蛋”，$B=$ “第二次取到红皮蛋”，则 $P(A)=\frac{3}{5}$，由于是有放回的抽取，所以$P(B)=\frac{3}{5}$.

$A\cap B=$ “两次都取到红皮蛋”，由于第一次取一个鸡蛋有 5 种取法，第二次取一个鸡蛋也有 5 种取法，于是两次共有 5×5 种取法，其中都取到红皮蛋的取法有 3×3 种，因此，两次都取到红皮蛋的概率为

$$P(A\cap B)=\frac{3\times3}{5\times5}=\frac{9}{25},$$

所以

$$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{3}{5}.$$

在例 1 中，事件 A 是否发生对事件 B 发生的概率没有影响，即

$$P(B|A)=P(B),$$

这时，我们称两个事件 A，B **相互独立**，并把这两个事件叫做**相互独立事件**.

在实际问题中，常常通过对事件本质进行分析就可知道它们是否相互独立，而不需要进行类似上面的计算去验证. 比如，依次抛掷两枚硬币，抛掷第一枚硬币的结果对第二枚硬币的结果没有影响. 又如，将一枚骰子连续抛掷 2 次，第一次抛得的点数对第二次抛得的点数也不会有影响，所以两次抛掷事件相互独立. 在例 1 中，我们还可通过计算得到

$$P(B|\overline{A})=P(B),$$

即第一次取到白皮蛋对第二次取到红皮蛋也没有影响.

一般地，当事件 A，B 相互独立时，A 与 $\overline{B}$，$\overline{A}$ 与 B，$\overline{A}$ 与 $\overline{B}$ 也相互独立.

由条件概率公式和相互独立事件 A，B 的定义，可以得到

$$P(B)=P(B|A)=\frac{P(A\cap B)}{P(A)},$$

即

$$P(A\cap B)=P(A)\times P(B). \qquad ③$$

这就是说，两个相互独立事件都发生的概率，等于每个事件发生的概率的积.

对于 n 个事件 A_1，A_2，…，A_n，如果其中任一个事件发生的概率不受其他事件是否发生的影响，则称 n 个事件 A_1，A_2，…，A_n 相互独立.

探索与研究

我们用一个例子来介绍 3 个事件相互独立所要满足的条件. 把抛掷硬币 A、B、C 得到正面的事件分别记作 A，B，C，如果有

$$P(A)=P(A|B)=P(A|C)=P(A|BC),$$
$$P(B)=P(B|C)=P(B|A)=P(B|AC),$$
$$P(C)=P(C|A)=P(C|B)=P(C|AB)$$

成立，我们就说这 3 个事件 A，B，C 相互独立.

这里，$P(A|B)$ 的含义是在硬币 B 掷出正面的条件下硬币 A 掷出正面的概率，它应该是 $\frac{1}{2}$；$P(A|BC)$ 的含义是在硬币 B、C 掷出正面的条件下硬币 A 掷出正面的概率，它也应该是 $\frac{1}{2}$；$P(A)$ 自然是 $\frac{1}{2}$……所以这些等式都是成立的. 于是我们可以说，事件 A，B，C 相互独立.

在实际问题中，对于 n 个事件，通常是考察这些事件的含义，用日常生活或生产中得到的经验来分析它们之间有没有影响，如果没有影响，或者影响可以忽略不计，就可以判断这 n 个事件是相互独立的.

如果事件 A_1，A_2，…，A_n 相互独立，那么这 n 个事件都发生的概率，等于每个事件发生的概率的积，即

$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)\times P(A_2)\times\cdots\times P(A_n), \qquad ③'$$

并且上式中任意多个事件 A_i 换成其对立事件后等式仍成立.

例 2 甲、乙两名篮球运动员分别进行一次投篮，如果两人投中的概率都是 0.6，计算：

（1）两人都投中的概率；

（2）其中恰有一人投中的概率；

（3）至少有一人投中的概率.

分析 甲、乙两人各投篮一次，甲（或乙）是否投中，对乙（或甲）投中的概率是没有影响的，也就是说，“甲投篮一次，投中”与“乙投篮一次，投中”是相互独立事件. 因此，可以求出这两个事件同时发生的概率. 同理可以分别求出，甲投中与乙未投中，甲未投中与乙投中，甲未投中与乙未投中同时发生的概率，从而可以得到所求的各个事件的概率.

解：（1）设 $A=$ “甲投篮一次，投中”，$B=$ “乙投篮一次，投中”，则 $A\cap B=$ “两人各投篮一次，都投中”. 由题意知，事件 A 与 B 相互独立，根据公式③所求概率为

$$P(A\cap B)=P(A)\cdot P(B)=0.6\times 0.6=0.36.$$

（2）事件“两人各投篮一次，恰好有一人投中”包括两种情况：一种是甲投中、乙未投中（事件 $A\cap\overline{B}$ 发生），另一种是甲未投中、乙投中（事件 $\overline{A}\cap B$ 发生）. 根据题意，这两种情况在各投篮一次时不可能同时发生，即事件 $A\cap\overline{B}$ 与 $\overline{A}\cap B$ 互斥，并且 A 与 $\overline{B}$，$\overline{A}$ 与 B 各自相互独立，因而所求概率为

$$\begin{aligned}P(A\cap\overline{B})+P(\overline{A}\cap B)&=P(A)\cdot P(\overline{B})+P(\overline{A})\cdot P(B)\\&=0.6\times(1-0.6)+(1-0.6)\times 0.6\\&=0.48.\end{aligned}$$

（3）事件“两人各投篮一次，至少有一人投中”的对立事件“两人各投篮一次，均未投中”的概率是

$$P(\overline{A}\cap\overline{B})=P(\overline{A})\cdot P(\overline{B})=(1-0.6)\times(1-0.6)=0.16.$$

因此，至少有一人投中的概率为

$$P(A \cup B)=1-P(\overline{A} \cap \overline{B})=1-0.16=0.84.$$

例 3 在一段线路中并联着三个独立自动控制的常开开关，只要其中有一个开关能够闭合，线路就能正常工作．假定在某段时间内每个开关能够闭合的概率都是 0.7，计算在这段时间内线路正常工作的概率．

分析 根据题意，这段时间内线路正常工作的概率，就是三个开关中至少有一个能闭合的概率，也就是三个开关都不能闭合的对立事件的概率，由于这段时间内三个开关是否能闭合相互之间没有影响，三个开关都不能闭合的概率可根据公式求出，从而可得到所求的概率．

解：分别记这段时间内开关 S_1、S_2、S_3 能够闭合为事件 A，B，C（图 2-3）．根据题意，$\overline{A}$，$\overline{B}$，$\overline{C}$ 相互独立．所以这段时间内至少有一个开关能够闭合，从而使线路能正常工作的概率是

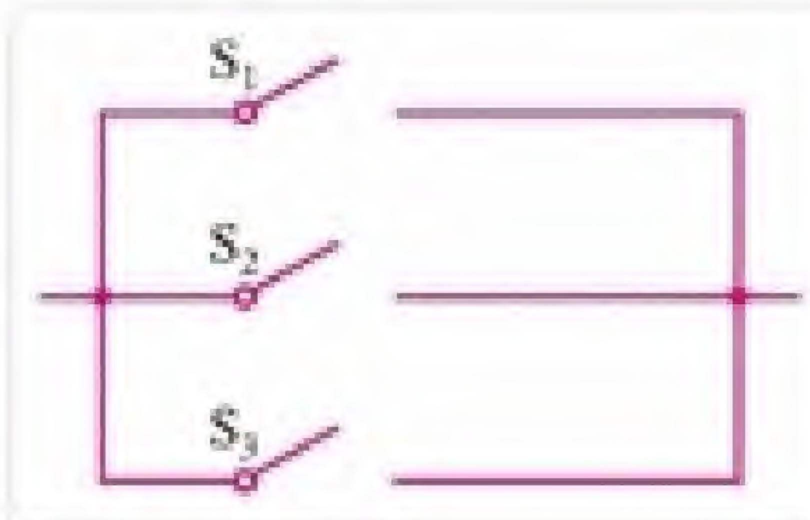


图 2-3

$$\begin{aligned}P(A \cup B \cup C)&=1-P(\overline{A} \cap \overline{B} \cap \overline{C})\\&=1-P(\overline{A})P(\overline{B})P(\overline{C})\\&=1-[1-P(A)][1-P(B)][1-P(C)]\\&=1-(1-0.7)(1-0.7)(1-0.7)\\&=1-0.3\times0.3\times0.3\\&=0.973.\end{aligned}$$

练习A

1. 对同一目标进行两次独立的射击，其命中的概率分别为 0.4 和 0.5，试求下列事件的概率：

   (1) 恰有一次命中；(2) 两次都命中．

2. 当开关 S_1 与 S_2 同时断开时电路断开．设 S_1，S_2 断开的概率分别为 0.5 和 0.7，且各开关相互独立．求电路为断开的概率．

3. 生产零件需要经过三道工序，在第一、二、三道工序中生产出废品的概率分别为 0.02，0.03，0.02，假设每道工序生产废品是独立事件．试求经过三道工序后得到的零件不是废品的概率．

4. 有一个问题，在半小时内，甲能解决它的概率是 $\frac{1}{2}$，乙能解决它的概率是 $\frac{1}{3}$，如果两人都试图独立地在半小时内解决它，计算：

   (1) 两人都未解决的概率；

   (2) 问题得到解决的概率．

练习B

1. 一个人的血型为O、A、B、AB型的概率分别为0.46，0.40，0.11，0.03，任意挑选5人，求下列事件的概率：
   (1) 两人为O型，其他三人分别为另外三种血型；
   (2) 三人为O型，两人为A型；
   (3) 没有一人为AB型.
2. 一个工人看管三台自动机床，在一小时内第一、二、三台机床不需要照顾的概率分别为0.9，0.8，0.85，在一小时的过程中，试求：
   (1) 没有一台机床需要照顾的概率；
   (2) 恰有两台机床需要照顾的概率；
   (3) 至少有一台机床需要照顾的概率；
   (4) 至少有两台机床需要照顾的概率.

2.2.3 独立重复试验与二项分布

本小节涉及的每次试验，只考虑有两个可能的结果 A 及 $\overline{A}$，并且事件 A 发生的概率相同. 在相同的条件下，重复地做 n 次试验，各次试验的结果相互独立，那么一般就称它们为**n 次独立重复试验**.

例如，对一批产品进行抽样检验，每次取一件，有放回地抽取 n 次，就是一个 n 次独立重复试验. 又如，某位篮球运动员进行 n 次投篮，如果每次投篮时的条件都相同，而且每次投中的概率相同，那么这也是一个 n 次独立重复试验. 在 n 次独立重复试验中，事件 A 恰好发生 $k(0\leqslant k\leqslant n)$ 次的概率问题叫做伯努利概型. 这是由于瑞士数学家雅各布·伯努利(1654—1705)对这方面的研究做了大量的工作.


雅各布·伯努利

本章章前语中提到，篮球运动员姚明在某一赛季罚篮命中率是80.9%. 我们把姚明在这一赛季罚篮的命中率当作他罚球得分的概率，则他每次罚球的得分服从 $p=0.809$ 的二点分布，即罚球一次得1分的可能性是0.809，得0分的可能性是0.191. 如果姚明在某场比赛中得到4次罚球机会，假设每次投篮都互不影响，那么他投中3次的可能性有多大呢?

如果用“⊙”代表投中，用“×”代表未投中，那么投球4次、投中3次有以下4种可能的情况(括号内为相应的概率)：

⊙⊙⊙×，$(0.809\times0.809\times0.809\times(1-0.809))$

⊙⊙×⊙，$(0.809\times0.809\times(1-0.809)\times0.809)$

⊙×⊙⊙，$(0.809\times(1-0.809)\times0.809\times0.809)$

×⊙⊙⊙，$((1-0.809)\times0.809\times0.809\times0.809)$

它们可以看成是从4个位置中任取3个填上“⊙”，最后的一个填上“×”，所有的取法为C_4^3种.

这就是说，在上面投篮4次、投中3次的4种情况中，每一种发生的概率都是

$$0.809^3\times(1-0.809)^{4-3}.$$

因为这4种情况彼此互斥，根据概率加法公式，投篮4次、恰好投中3次的概率为

$$\begin{aligned}&P(\text{“⊙⊙⊙×”})+P(\text{“⊙⊙×⊙”})+P(\text{“⊙×⊙⊙”})+P(\text{“×⊙⊙⊙”})\\&=C_4^3\times0.809^3\times(1-0.809)^{4-3}\\&=4\times0.809^3\times0.191=0.405.\end{aligned}$$

也就是说，姚明罚球4投3中的概率还不到0.5，这个结果与我们的感觉可能有些差距. 实际上，还有别的结果尚未计算概率. 类似地还可算出4投4中的概率为0.428，于是，可以看出姚明4次罚球投中3次以上的概率很大，为$0.405+0.428$，即0.833.

在上面的例子中，4次投篮是4次独立重复试验，也可以看成是进行4次独立的二点分布试验.

一般地，事件A在n次试验中发生k次，共有C_n^k种情形，由试验的独立性知A在k次试验中发生，而在其余$n-k$次试验中不发生的概率都是$p^k(1-p)^{n-k}$，所以由概率加法公式知，如果在一次试验中事件A发生的概率是p，那么在n次独立重复试验中，事件A恰好发生k次的概率为

$$P_n(k)=C_n^kp^k(1-p)^{n-k}\ (k=0,1,2,\cdots,n). \quad ④$$

例1 在人寿保险事业中，很重视某一年龄段的投保人的死亡率，假如每个投保人能活到65岁的概率为0.6，试问3个投保人中：

(1) 全部活到65岁的概率；

(2) 有2个活到65岁的概率；

(3) 有1个活到65岁的概率；

(4) 都活不到65岁的概率.

解：设$A=$“1个投保人能活到65岁”，则$\overline{A}=$“1个投保人活不到65岁”.

$P(A)=p=0.6$，

$P(\overline{A})=1-p=1-0.6=0.4$.

3个投保人活到65岁的人数相当于作3次独立重复试验中事件A发生的次数，由公式④有

(1) $P_3(3)=C_3^3\cdot0.6^3\cdot(1-0.6)^0=0.216$；

(2) $P_3(2)=C_3^2\cdot0.6^2\cdot(1-0.6)^1=0.432$；

(3) $P_3(1)=C_3^1\cdot0.6^1\cdot(1-0.6)^2=0.288$；

(4) $P_3(0)=C_3^0\cdot0.6^0\cdot(1-0.6)^3=0.064$.

在公式④中，若将事件 A 发生的次数设为 X，事件 A 不发生的概率为 $q=1-p$，那么在 n 次独立重复试验中，事件 A 恰好发生 k 次的概率是

$$P(X=k)=C_n^k p^k q^{n-k},$$

其中 $k=0, 1, 2, \cdots, n$. 于是得到 X 的分布列

X	0	1	…	k	…	n
P	$C_n^0p^0q^n$	$C_n^1p^1q^{n-1}$	…	$C_n^kp^kq^{n-k}$	…	$C_n^np^nq^0$

由于表中的第二行恰好是二项式展开式

$$(q+p)^n=C_n^0p^0q^n+C_n^1p^1q^{n-1}+\cdots+C_n^kp^kq^{n-k}+\cdots+C_n^np^nq^0$$

各对应项的值，所以称这样的离散型随机变量 X 服从参数为 n，p 的**二项分布**，记作

$$X\sim B(n, p).$$

例 2 100 件产品中有 3 件不合格品，每次取一件，有放回地抽取三次，求取得不合格品件数 X 的分布列.

解：X 可能取的值为 0，1，2，3. 由于是有放回地每次取一件，连续取三次，所以这相当于作 3 次独立重复试验，一次抽取到不合格品的概率 $p=0.03$. 因此

$$P(X=0)=C_3^0\cdot 0.03^0\cdot(1-0.03)^3=0.912\ 673,$$
$$P(X=1)=C_3^1\cdot 0.03^1\cdot(1-0.03)^2=0.084\ 681,$$
$$P(X=2)=C_3^2\cdot 0.03^2\cdot(1-0.03)^1=0.002\ 619,$$
$$P(X=3)=C_3^3\cdot 0.03^3\cdot(1-0.03)^0=0.000\ 027.$$

分布列为

X	0	1	2	3
P	0.912 673	0.084 681	0.002 619	0.000 027

例 3 将一枚均匀硬币随机掷 100 次，求正好出现 50 次正面的概率.

解：掷一次硬币可以看作一次试验，每次有两个可能的结果：出现正面或不出现正面. 由于硬币是均匀的，所以出现正面的概率为 0.5，因此掷 100 次硬币可以看作 100 次独立重复试验. 如果用 X 表示出现正面的次数，则 X 服从 $n=100$，$p=0.5$ 的二项分布，那么所求概率为

$$P(X=50)=C_{100}^{50}p^{50}(1-p)^{100-50}=C_{100}^{50}\times 0.5^{50}\times 0.5^{50}\approx 0.08.$$

练习A

1. 某班有 50 个学生，假设每个学生早上到校时间相互没有影响，并且迟到的概率均为 0.05，试求这个班某天正好有 4 个学生迟到的概率.
2. 某学生在最近的 15 次数学测验中有 5 次不及格. 按照这个成绩，他在接下来的 10 次测验中(1) 全及格；(2) 全不及格；(3) 恰好 5 次及格的概率各是多少？

3. 一次测量中出现正误差和负误差的概率都是0.5，在3次测量中，恰好出现2次正误差的概率是多少？恰好出现2次负误差的概率是多少？

4. 某射手射击5次，每次命中的概率为0.6，求下列事件的概率：

(1) 5次中有3次中靶；

(2) 5次中至少有3次中靶.

5. 已知某种疗法的治愈率是90%，在对10位病人采用这种疗法后，正好有9人被治愈的概率是多少？

练习B

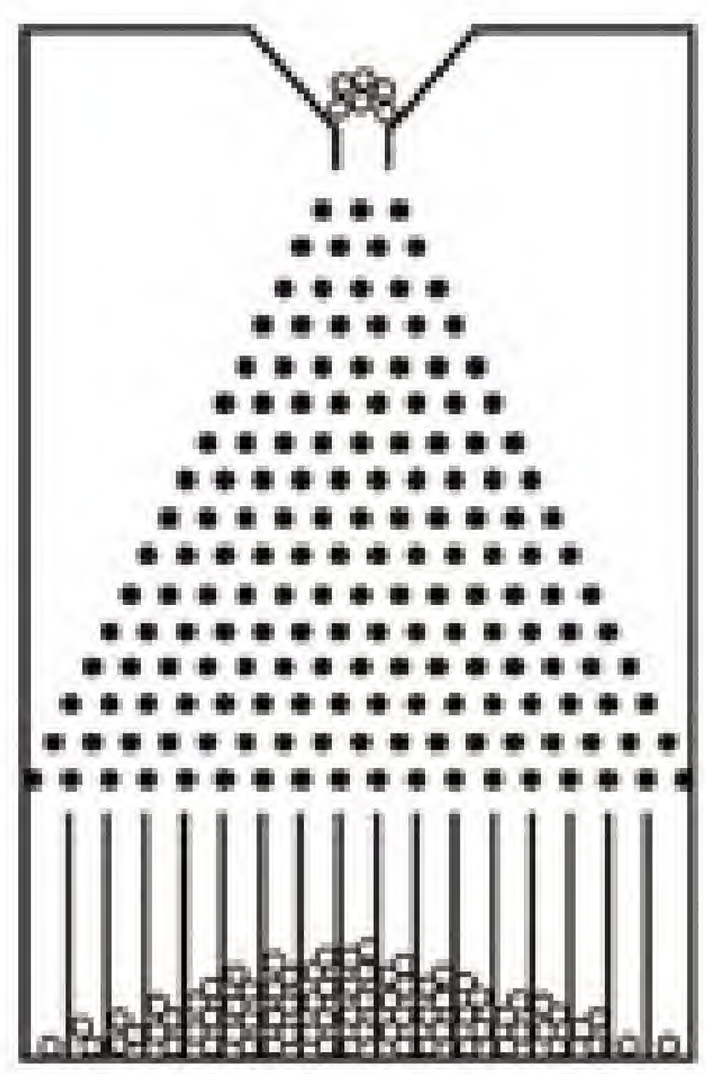
1. 假定人在一年365天中的任一天出生的概率是一样的，某班级有50名同学，其中有两个以上的同学生于元旦的概率是多少？

2. 设顾客需要27号鞋的概率为0.2，求鞋店上午开门营业后，头5名顾客中：

(1) 有1人要买27号鞋的概率；

(2) 至少有1人要买27号鞋的概率.

探索与研究



高尔顿(钉)板是在一块竖起的木板上钉上一排排互相平行、水平间隔相等的铁钉(如右图所示)，并且每一排钉子数目都比上一排多一个，一排中各个钉子正好对准上面一排两个相邻铁钉的正中央. 从入口处放入一个直径略小于两颗钉子间隔的小球，当小球从两钉之间的间隙下落时，由于碰到下一排铁钉，它将以相等的可能性向左或向右落下，接着小球再通过两钉的间隙，又碰到下一排铁钉. 如此继续下去，小球最后落入下方条状的格子内.

有兴趣的同学可以通过以下的问题研究高尔顿板与二项分布的关系.

1. 通过高尔顿板实验课件，做1 000个小球的高尔顿板试验，看看小球在格子中的分布形状是怎样的？

2. 计算小球落入各个格子所有可能路线的数目.（提示：考虑它与杨辉三角的关系）

3. 计算小球落入各个格子的概率.

根据上面这些问题的结果你能得出什么结论?

习题 2-2 A

1. 若 10 件产品中包含 2 件废品，今在其中任取两件，求：

(1) 取出的两件中至少有一件是废品的概率；

(2) 已知取出的两件中有一件是废品的条件下，另一件也是废品的概率；

(3) 已知两件中有一件不是废品的条件下，另一件是废品的概率.

2. 设一个班级中有$\frac{1}{3}$的女生，$\frac{1}{5}$的三好学生，而三好学生中女生占$\frac{1}{3}$，若从此班级中任选一名代表参加夏令营活动，试问在已知没有选上女生的条件下，选的是一位三好学生的概率是多少?

3. 已知一批玉米种子的出苗率为 0.9，现每穴种两粒，问一粒出苗一粒不出苗的概率是多少?

4. 在某个学校里，所有学生都学习数学和英语，随机找出一个学生，他数学不及格的概率是 0.15，英语不及格的概率是 0.05，这两门都不及格的概率是 0.04，问：

(1) 数学不及格和英语不及格这两个事件是相互独立的吗?

(2) 已知一个学生英语不及格，他数学不及格的概率是多少?

(3) 已知一个学生数学不及格，他英语不及格的概率是多少?

5. 某棒球手一次击球得 1 分的概率平均为 0.2，在 5 次击球中他得 2 分的概率是多少?

6. 某气象站天气预报的准确率为 80%，计算(结果保留两个有效数字)：

(1) 5 次预报中恰有 4 次准确的概率；

(2) 5 次预报中至少有 4 次准确的概率.

习题 2-2 B

1. 盒子里装有 16 个球，其中 6 个是玻璃球，10 个是木质球，玻璃球中有 2 个是红色的，4 个是蓝色的，木质球中有 3 个是红色的，7 个是蓝色的，现从中任取一个发现是蓝球，问该球是玻璃球的概率是多少?

2. 在某售楼中心，最近的 100 位顾客中有一位买了某房产商出售的住房. 根据这一比例，试问在接下来的 50 位顾客中(1) 恰好一位；(2) 至少一位；(3) 多于一位顾客买这个房产商的房子的概率各是多少?

2.3 随机变量的数字特征

由超几何分布和二项分布等离散型随机变量的分布，我们知道离散型随机变量的分布列能够完全描述随机变量取值的概率规律. 但是，在许多实际问题中，还需要了解离散型随机变量的某种特征，例如离散型随机变量的平均取值大小和取值的集中程度. 我们把这种反映概率分布的某种特征的数值，叫做离散型随机变量的数字特征. 下面我们来介绍两个最基本的离散型随机变量的数字特征.

2.3.1 离散型随机变量的数学期望

某学校为了解交通拥堵对同学们上学迟到的影响情况，每天记录由于交通问题迟到的同学人数. 下表是在 100 天中每天由于交通原因迟到人数的情况：

人数	0	1	2	3
天数	30	30	20	20

那么这所学校每天平均有多少人由于交通原因迟到呢？通过计算可得 100 天中记录的迟到次数的总和是

$$0\times30+1\times30+2\times20+3\times20=130,$$

那么平均每天迟到人数为

$$\frac{0\times30+1\times30+2\times20+3\times20}{100}=\frac{130}{100}=1.3.$$

上式可改写成

$$\begin{aligned}&0\times\frac{30}{100}+1\times\frac{30}{100}+2\times\frac{20}{100}+3\times\frac{20}{100}\\&=0\times0.3+1\times0.3+2\times0.2+3\times0.2\\&=1.3.\end{aligned}$$

这种方法是用迟到人数乘以各自的频率，然后求和得到的. 我们可把迟到人数和对应的频率列表为

迟到人数	0	1	2	3
频率	0.3	0.3	0.2	0.2

由频率与概率的关系可知，概率可以理解为频率的稳定值，所以如果这所学校在一天中由于交通原因迟到人数为 X，则 X 的概率分布为

X	0	1	2	3
P	0.3	0.3	0.2	0.2

那么平均每天迟到人数为

$$0\times0.3+1\times0.3+2\times0.2+3\times0.2=1.3\ (\text{人}).$$

一般地，设一个离散型随机变量 X 所有可能取的值是 x_1，x_2，…，x_n，这些值对应的概率是 p_1，p_2，…，p_n，则

$$E(X)=x_1p_1+x_2p_2+\cdots+x_np_n$$

叫做这个**离散型随机变量 X 的均值**或**数学期望**(简称**期望**).

离散型随机变量的数学期望刻画了这个离散型随机变量的平均取值水平.

思考：怎样理解“平均每天迟到人数为 1.3 人”这句话的意义.

由数学期望的定义可以知道，若随机变量 X 服从参数为 p 的二点分布，则

$$E(X)=1\times p+0\times(1-p)=p,$$

这表明在一次二点分布试验中，离散型随机变量 X 的期望取值为 p.

例如，某篮球运动员罚篮命中率是 0.7，他平均说来一次罚篮期望得到的分数就是 0.7 分. 那么平均来看，他 10 次罚篮能够期望得到多少分呢？我们猜想他会得到 $10\times0.7=7$分，下面来证明这一猜想.

设离散型随机变量 X 服从参数为 n 和 p 的二项分布，由 X 的分布列

$$P(X=k)=C_n^kp^kq^{n-k},\ k=0,\ 1,\ 2,\ \cdots,\ n$$

和数学期望的定义式得到

$$\begin{aligned}E(X)&=0\times C_n^0p^0q^n+1\times C_n^1p^1q^{n-1}+2\times C_n^2p^2q^{n-2}+\cdots+\\&\quad k\times C_n^kp^kq^{n-k}+\cdots+n\times C_n^np^nq^0\\&=np(C_{n-1}^0p^0q^{n-1}+C_{n-1}^1p^1q^{n-2}+\cdots+\\&\quad C_{n-1}^{k-1}p^{k-1}q^{(n-1)-(k-1)}+\cdots+C_{n-1}^{n-1}p^{n-1}q^0)❶\\&=np(p+q)^{n-1}=np.\end{aligned}$$

注

❶ 这是由于 $kC_n^k=nC_{n-1}^{k-1}$，证明见本章附录.

所以，

$$\boxed{E(X)=np.}$$

由于上面提到的篮球运动员 10 次罚篮得到的分数 $X\sim B(10,\ 0.7)$，所以 10 次罚篮能够期望得到的分数 $E(X)=np=10\times0.7=7$ (分).

若离散型随机变量 X 服从参数为 N，M，n 的超几何分布，则

$$\boxed{E(X)=\frac{nM}{N}.}❷$$

注

❷ 证明过程见本章附录.

例 1 根据历次比赛或训练记录，甲、乙两射手在同样的条件下进行射击，成绩的分布列如下：

射手	8 环	9 环	10 环
甲	0.3	0.1	0.6
乙	0.2	0.5	0.3

试比较甲、乙两射手射击水平的高低.

解：设甲、乙两射手射击一次所得的环数分别为 X_1，X_2，则

$$E(X_1)=8\times0.3+9\times0.1+10\times0.6=9.3,$$
$$E(X_2)=8\times0.2+9\times0.5+10\times0.3=9.1.$$

这就是说射手甲射击所得环数的数学期望比射手乙射击所得环数的数学期望高，从而说明甲的平均射击水平比乙的稍高一点．如果两人进行比赛，甲赢的可能性较大.

例 2 一个袋子里装有大小相同的 5 个白球和 5 个黑球，从中任取 4 个，求其中所含白球个数的期望.

解：根据题目知所含白球数 X 服从参数 $N=10$，$M=5$，$n=4$ 的超几何分布，则

$$E(X)=\frac{nM}{N}=\frac{4\times5}{10}=2.$$

所以从中任取 4 个球平均来说会含有 2 个白球.

例 3 根据气象预报，某地区下个月有小洪水的概率为 0.25，有大洪水的概率为 0.01．设工地上有一台大型设备，为保护设备有以下三种方案.

方案 1：运走设备，此时需花费 3 800 元.

方案 2：建一保护围墙，需花费 2 000 元．但围墙无法防止大洪水，当大洪水来临，设备受损，损失费为60 000 元.

方案 3：不采取措施，希望不发生洪水．此时大洪水来临损失60 000元，小洪水来临损失10 000元.

试比较哪一种方案好.

解：关键是看哪种方案的花费与期望损失和最小.

对于方案 1，花费为 3 800 元，损失为 0 元，花费与期望损失之和为 3 800 元；

对于方案 2，花费为 2 000 元，损失费的分布列为

损失费（元）	60 000	0
概率	0.01	0.99

期望损失为 $60\ 000\times0.01+0\times0.99=600$（元），所以花费与期望损失之和为 $2\ 000+600=2\ 600$（元）；

对于方案 3，花费为 0 元，损失费的分布列为

损失费（元）	60 000	10 000	0
概率	0.01	0.25	0.74

期望损失为 $60\ 000\times0.01+10\ 000\times0.25+0\times0.74=3\ 100$（元），所以花费与期望损失之和为 3 100 元.

比较三种方案，我们发现第二种方案的花费与期望损失之和最小，故方案 2 较好.

2.3.2 离散型随机变量的方差

在实际问题中，我们除了要知道一个离散型随机变量取值的平均水平，还经常要知道离散型随机变量取值波动的大小.

我们知道，样本方差可以表示样本数据的波动情况．例如，10 个学生在一次数学考试中的成绩如下表所示：

分数	100	90	80	70
人数	2	4	3	1

对于由这 10 个学生的成绩组成的样本来说，可以算出样本的平均数是 87，因而样本方差是

$$\frac{1}{10}[2\times(100-87)^2+4\times(90-87)^2+3\times(80-87)^2+(70-87)^2]$$

$$=(100-87)^2\times\frac{2}{10}+(90-87)^2\times\frac{4}{10}+(80-87)^2\times\frac{3}{10}+(70-87)^2\times\frac{1}{10}.$$

从随机变量的角度来看，如果从这 10 个学生的成绩中任取一个学生的成绩，把它看作是离散型随机变量 X，那么 X 取值 100，90，80，70 的概率分别是 $\frac{2}{10}$，$\frac{4}{10}$，$\frac{3}{10}$，$\frac{1}{10}$，数学期望值 $E(X)$ 是 87，则上式就描述了离散型随机变量 X 围绕 $E(X)$ 波动的平均大小.

一般地，设一个离散型随机变量 X 所有可能取的值是 x_1，x_2，…，x_n，这些值对应的概率是 p_1，p_2，…，p_n，则

$$D(X)=(x_1-E(X))^2p_1+(x_2-E(X))^2p_2+\cdots+(x_n-E(X))^2p_n$$

叫做这个**离散型随机变量 X 的方差**.

离散型随机变量的方差反映了离散型随机变量取值相对于期望的平均波动大小(或说离散程度).

$D(X)$ 的算术平方根 $\sqrt{D(X)}$ 叫做**离散型随机变量 X 的标准差**，它也是一个衡量离散型随机变量波动大小的量.

例 1 甲、乙两名射手在同一条件下进行射击，分布列如下：

射手甲：

所得环数 X_1	10	9	8
概率 P	0.2	0.6	0.2

射手乙：

所得环数 X_2	10	9	8
概率 P	0.4	0.2	0.4

谁的射击水平比较稳定？

解：$E(X_1)=10\times0.2+9\times0.6+8\times0.2=9$，

$$D(X_1)=(10-9)^2\times0.2+(9-9)^2\times0.6+(8-9)^2\times0.2$$
$$=0.2+0.2=0.4，$$

$$E(X_2)=10\times0.4+9\times0.2+8\times0.4=9，$$

$$D(X_2)=(10-9)^2\times0.4+(9-9)^2\times0.2+(8-9)^2\times0.4$$
$$=0.4+0.4=0.8.$$

由 $D(X_1)<D(X_2)$ 可知，甲的射击水平比乙稳定.

例 2 已知某离散型随机变量 X 服从的分布列为

X	1	0
P	p	q

且 $0<p<1$，$q=1-p$，求 $D(X)$.

解：由题目知 X 服从二点分布. 所以

$$E(X)=p，$$

$$D(X)=(1-p)^2\cdot p+(0-p)^2\cdot q=q^2p+p^2q=pq.$$

这表明在二点分布试验中，离散型随机变量 X 围绕期望的平均波动大小为 pq.

现在考虑这样一个问题：设某射击选手射击的命中率是 0.8，那么他 10 次独立射击时的命中次数平均波动大小是多少呢？我们有下面的结论.

若离散型随机变量 X 服从参数为 n 和 p 的二项分布，则

$$\boxed{D(X)=npq\quad(q=1-p).}$$ ❶

注

❶ 证明过程见本章附录.

所以这位射击选手 10 次独立射击命中次数相对于期望的平均波动大小为 $10\times0.8\times0.2=1.6$.

例 3 已知某离散型随机变量 X 服从下面的二项分布：

$$P(X=k)=C_4^k0.1^k0.9^{4-k}\quad(k=0,1,2,3,4)，$$

求 $E(X)$ 和 $D(X)$.

解：根据题目知道离散型随机变量 X 服从参数 $n=4$ 和 $p=0.1$ 的二项分布，所以

$$E(X)=np=4\times0.1=0.4，$$

$$D(X)=npq=4\times0.1\times0.9=0.36.$$

习题 2-3 A

1. 设离散型随机变量 X 的分布列为

X	0	1	2	3	4	5
P	0.1	0.2	0.3	0.2	0.1	0.1

求 $E(X)$.

2. 两台生产同一零件的车床，设一天生产中次品的分布列分别为

甲	0	1	2	3
P	0.4	0.3	0.2	0.1

乙	0	1	2	3
P	0.3	0.5	0.2	0

如果两台车床在一天中的产量相同，试问哪台车床期望的次品少？

3. 已知某彩票中心发行彩票，每 100 000 张设一个奖，奖金为 10 000 元. 某人购买一张彩票，问这个人能期望得到多少奖金？

4. 设离散型随机变量 X 的分布列为

X	-2	-1	0	1	2
P	0.2	0.1	0.1	0.4	0.2

求 $D(X)$.

5. 掷一个骰子所得的点数为 X，求 $D(X)$.

6. 班上有 45 名同学，其中 30 名男生、15 名女生，老师随机地抽查了 5 名同学的作业，用 X 表示抽查到的女生的人数，求 $E(X)$.

7. 从装有 3 个白球和 2 个黑球的布袋中摸取一球，有放回的摸取 5 次，求摸得的白球数 X 的数学期望与方差.

习题 2-3 B

1. 某小组几名同学做抛掷两枚硬币游戏，抛前先猜测结果，得分是：猜对出现两个正面(或两个反面)得 10 分，猜错扣 1 分；猜对一正一反得 6 分，猜错扣 2 分. 得分高者获胜. 甲同学在抛前应如何猜测，才能使他获胜的期望较高.

2. 某彩票中心发行彩票 10 万张，每张 1 元. 设一等奖 1 个，奖金 1 万元；二等奖 2 个，奖金各 5 千元；三等奖 10 个，奖金各 1 千元；四等奖 100 个，奖金各 1 百元；五等奖1 000个，奖金各 10 元. 试求每张彩票的期望获利金额是多少？

2.4 正态分布

看下面的例子.

某钢铁加工厂生产内径为 25.40 mm 的钢管，为了检验产品的质量，从一批产品中任取 100 件检测，测得它们的实际尺寸如下：

25.39	25.36	25.34	25.42	25.45	25.38	25.39
25.42	25.47	25.35	25.41	25.43	25.44	25.48
25.45	25.43	25.46	25.40	25.51	25.45	25.40
25.39	25.41	25.36	25.38	25.31	25.56	25.43
25.40	25.38	25.37	25.44	25.33	25.46	25.40
25.49	25.34	25.42	25.50	25.37	25.35	25.32
25.45	25.40	25.27	25.43	25.54	25.39	25.45
25.43	25.40	25.43	25.44	25.41	25.53	25.37
25.38	25.24	25.44	25.40	25.36	25.42	25.39
25.46	25.38	25.35	25.31	25.34	25.40	25.36
25.41	25.32	25.38	25.42	25.40	25.33	25.37
25.41	25.49	25.35	25.47	25.34	25.30	25.39
25.36	25.46	25.29	25.40	25.37	25.33	25.40
25.35	25.41	25.37	25.47	25.39	25.42	25.47
25.38	25.39					

把这批产品的内径尺寸看作一个总体，那么这 100 件产品的实际尺寸就是一个容量为 100 的样本．由数学 3 中 2.2.1 节知识可得到这组样本数据的频率分布直方图(图 2-4).

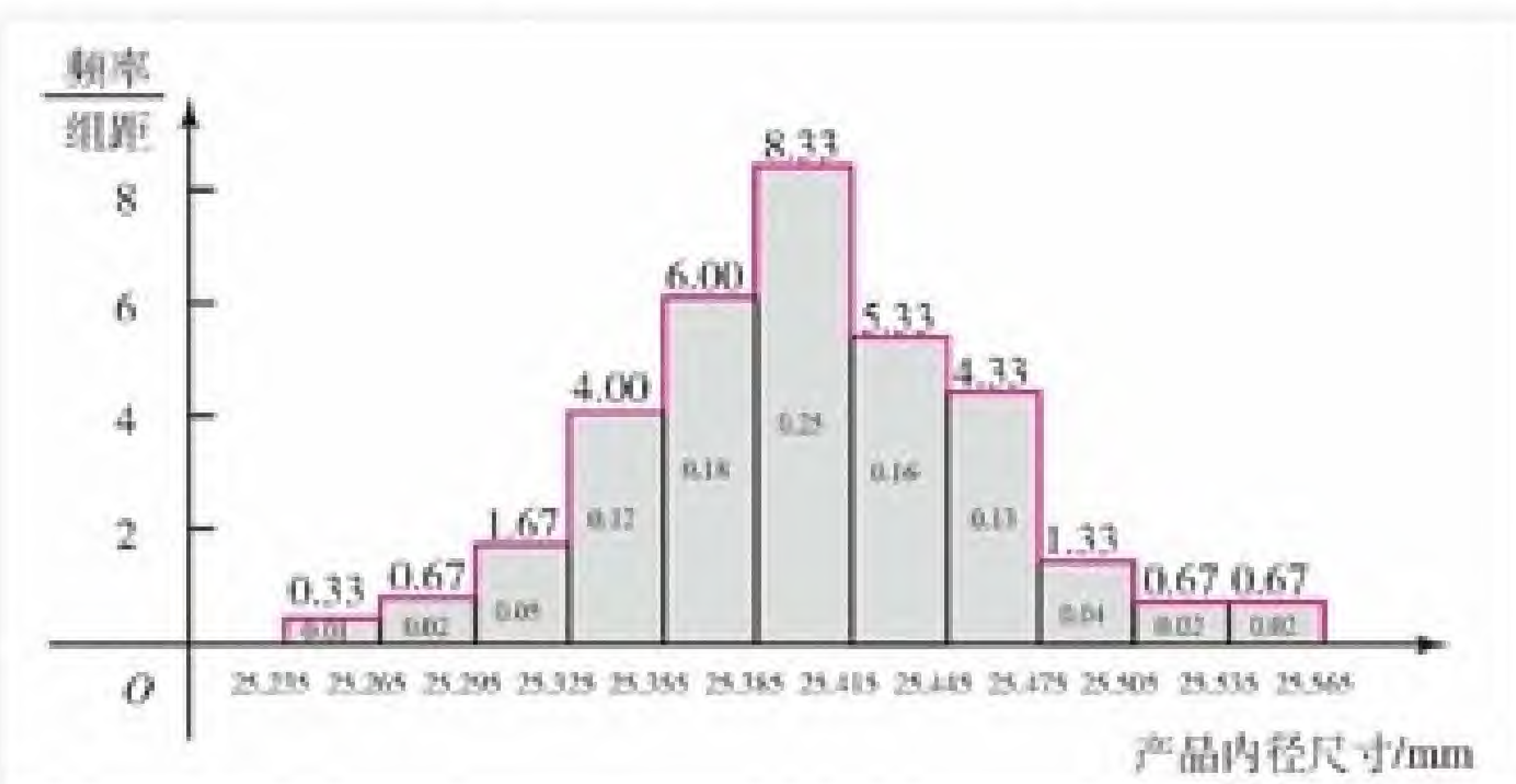


图 2-4

当样本容量 n 越来越大时，分组越来越细．频率直方图上面的折线越接近于图 2-5 中的曲线.

从随机变量的角度来看，如果把样本中的任一个产品尺寸看作随机变量 X，则这条曲线通常称为 X 的**概率密度曲线**. 这条曲线位于横轴的上方，它与横轴一起所围成的面积是 1，而随机变量 X 落在指定的两个数 a，b 之间的概率，就是图中带斜线部分的面积. 在我们这里就是指产品尺寸落在区间 (a, b) 内的概率，由于 a，b 是在产品尺寸范围内任意取值的，所以这条概率曲线就能精确地反映 X 取值的规律. 概率密度曲线反映变化规律所起的作用与离散型随机变量分布列的作用是相同的.

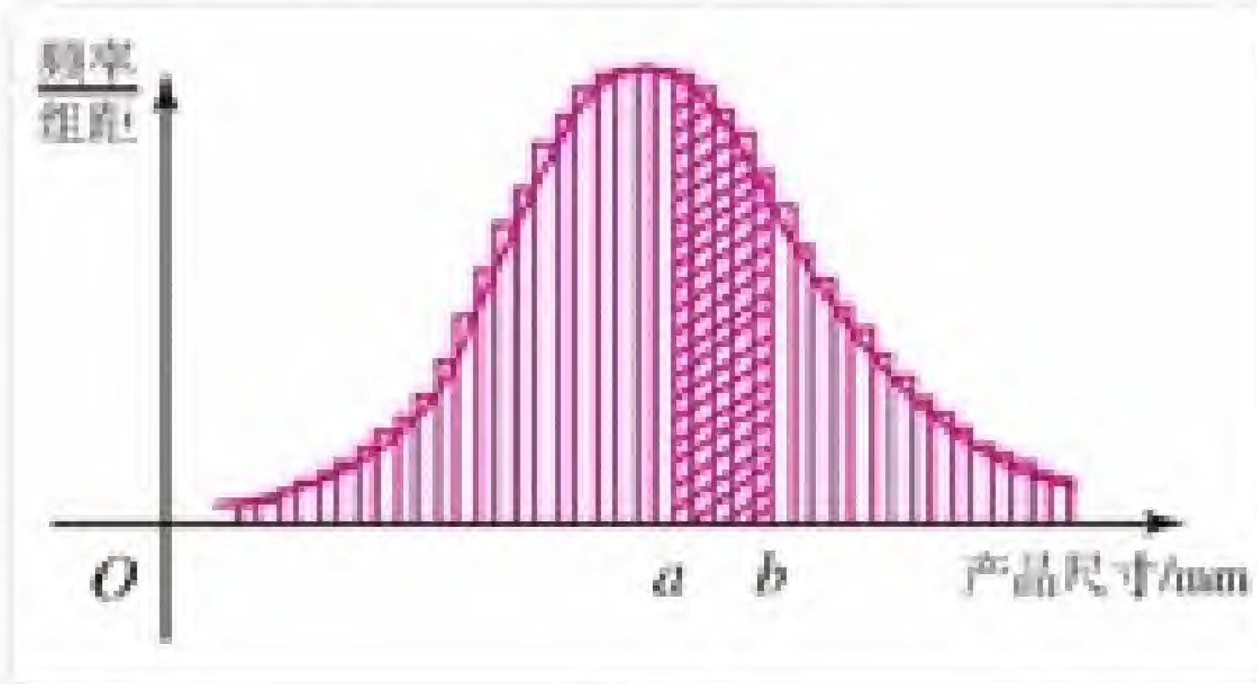


图 2-5

在生产、科研和日常生活中，经常会遇到这样一类随机现象，它们是由一些互相独立的偶然因素所引起的，而每一个这种偶然因素在总体的变化中都只是起着均匀、微小的作用. 例如，钢铁加工厂生产钢管时，加工零件的机器的磨损程度、使用的材料的差异、工人操作的习惯、周围的环境的温度等因素都可能会对钢管内径的尺寸起微小的影响，导致产品内径尺寸的波动. 表示这类随机现象的随机变量的概率分布一般近似服从**正态分布**. 服从正态分布的随机变量叫做**正态随机变量**，简称**正态变量**. 正态分布是自然界中最常见的一种分布，在理论研究和实际应用中都有非常重要的作用.

正态变量概率密度曲线的函数表达式为

$$f(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad x\in\mathbf{R} \qquad ⑤$$

其中 μ，σ 是参数，且 $\sigma>0$，$-\infty<\mu<+\infty$.

式⑤中的参数 μ 和 σ 分别为正态变量的数学期望和标准差. 期望为 μ、标准差为 σ 的正态分布通常记作 $N(\mu, \sigma^2)$. 正态变量的概率密度函数的图象叫做**正态曲线**. 图 2-6 画出了参数 μ 和 σ 取不同值的正态曲线. 其中，(1) $\mu=-1$，$\sigma=0.5$，(2) $\mu=1$，$\sigma=0.5$，(3) μ 都等于 0，而参数 σ 分别等于 0.5，1，2. 我们把数学期望为 0，标准差为 1 的正态分布叫做标准正态分布.

从图 2-6 可以看出，正态曲线具有以下性质：

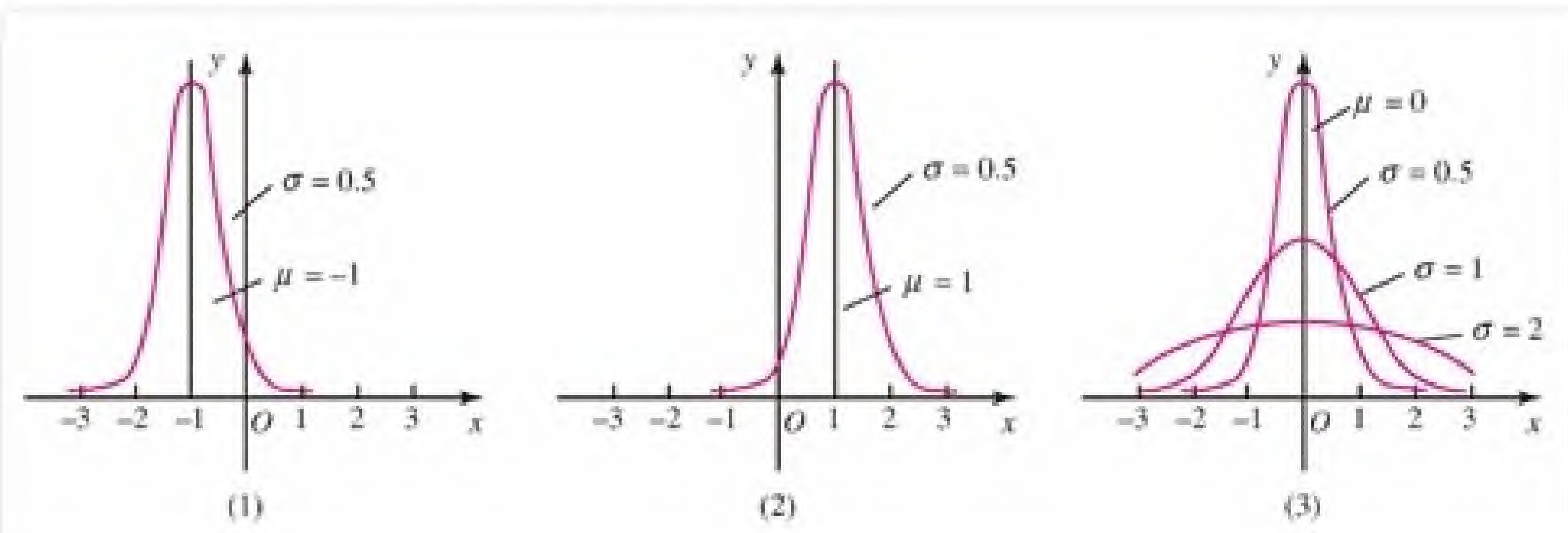


图 2-6

(1) 曲线在 x 轴的上方，并且关于直线 $x=\mu$ 对称；

(2) 曲线在 $x=\mu$ 时处于最高点，并由此处向左右两边延伸时，曲线逐渐降低，呈现“中间高，两边低”的形状；

(3) 曲线的形状由参数 σ 确定，σ 越大，曲线越“矮胖”；σ 越小，曲线越“高瘦”.

从理论上可以证明，正态变量在区间 $(\mu-\sigma,\ \mu+\sigma)$，$(\mu-2\sigma,\ \mu+2\sigma)$，$(\mu-3\sigma,\ \mu+3\sigma)$ 内，取值的概率分别是 68.3%，95.4%，99.7%，如图 2-7 所示.

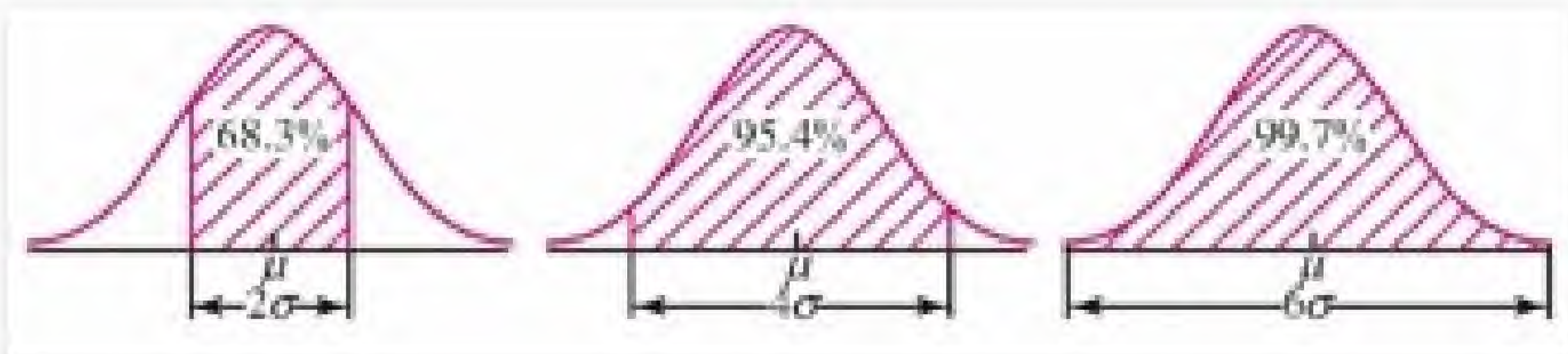


图 2-7

例如，当 $\mu=0$，$\sigma=1$ 时，正态变量（这时称它为标准正态变量）在区间 $(-1,\ 1)$，$(-2,\ 2)$，$(-3,\ 3)$ 内取值的概率分别是 68.3%，95.4%，99.7%. 若已知本节开头的钢铁加工厂生产的钢管内径尺寸 $X\sim N(25.40,\ 0.05^2)$，例如，对于该厂的 1 000 个产品，则利用正态分布可估计出其中产品内径尺寸在 $25.40-0.05\sim25.40+0.05$ 范围内的产品个数约为 683，在 $25.40-2\times0.05\sim25.40+2\times0.05$ 范围内的产品个数约为 954，在 $25.40-3\times0.05\sim25.40+3\times0.05$ 范围内的产品个数约为 997.

由于正态变量在 $(-\infty,\ +\infty)$ 内取值的概率是 1，由上所述，容易推出，它在区间 $(\mu-2\sigma,\ \mu+2\sigma)$ 之外取值的概率是 4.6%，在区间 $(\mu-3\sigma,\ \mu+3\sigma)$ 之外取值的概率是 0.3%. 于是，正态变量的取值几乎都在距 $x=\mu$ 三倍标准差之内，这就是正态分布的 3σ 原则.

例如，在工业生产中，工人操作车床在稳定的状态下加工出的零件尺寸一般服从正态分布，我们可以用正态分布的 3σ 原则来对加工出来的零件进行质量控制，在下面的 3.2 节中，还要讨论这一问题.

习题 2-4 A

1. 已知某厂生产的某种型号卡车轮胎的使用寿命（单位：km）服从正态分布 $N(36\,203,\ 4\,827^2)$. 一汽车公司一次从此厂买了 500 个轮胎，利用正态分布估计使用寿命分别在：

   (1) $36\,203-2\times4\,827\sim36\,203+2\times4\,827$；

   (2) $36\,203-3\times4\,827\sim36\,203+3\times4\,827$

   范围内的轮胎个数.

2. 某糖厂用自动打包机打包，每包重量 X(kg) 服从正态分布 $N(100,\ 1.2^2)$. 一公司从该糖厂进货 1 500 包，试估计重量在下列范围内的糖包数量：

   (1) $(100-1.2,\ 100+1.2)$；

（2）$(100-3\times1.2,\ 100+3\times1.2)$.

习题 2-4 B

在医疗卫生工作中，经常用到人的各种生理生化指标（如身高、红细胞数、血糖浓度等）的正常值．所谓正常值是指正常人的各种生理生化指标的观察值．由于生物的变异和环境条件不同，这些值是有波动的．如果已知正常人的某项生化指标服从正态分布，可根据大量的调查资料求 μ 和 σ，然后把在 $(\mu-2\sigma,\ \mu+2\sigma)$（或 $(\mu-3\sigma,\ \mu+3\sigma)$）区间的观察值作为 95.4%（或 99.7%）的正常值范围．已知某地区 12 岁男孩身高服从正态分布，已求出 $\mu=143.1$，$\sigma=5.668$．求该地 12 岁男孩身高的 95.4% 和 99.7% 正常值范围．

计算机上的练习

使用计算机作图软件，根据正态分布的概率密度函数

$$f(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in\mathbf{R}$$

画出

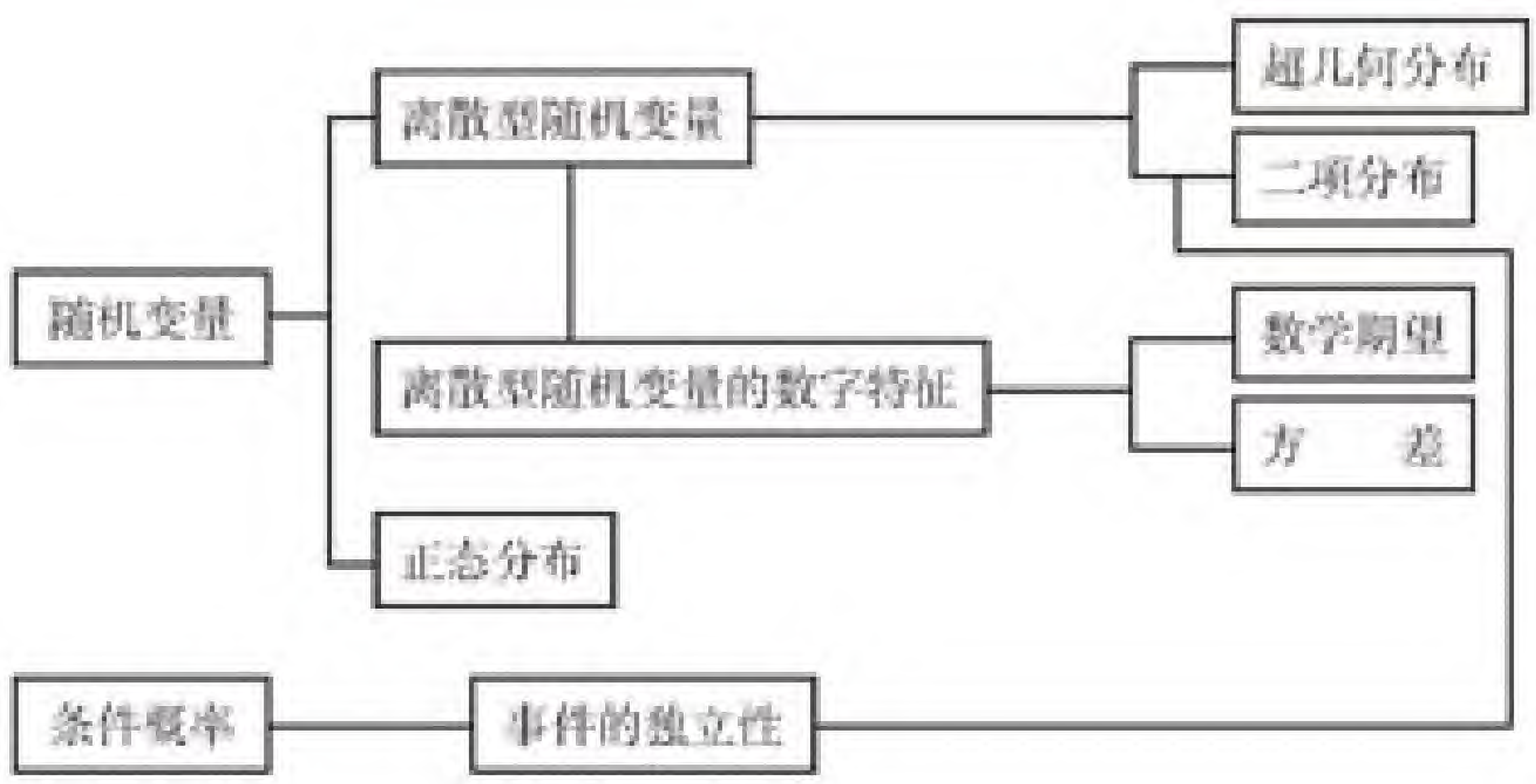
（1）参数 $\mu=0$，参数 σ 分别等于 0.5，1，2；

（2）参数 $\sigma=1$，参数 μ 分别等于 -1，1，2

的正态曲线，说出参数 μ 与 σ 对曲线形状位置的影响．

本章小结

Ⅰ 知识结构



Ⅱ 思考与交流

1. 什么叫随机变量？你能举一个生活中的离散型随机变量的例子吗？
2. 什么是条件概率？如何计算 $P(B|A)$.
3. 什么是相互独立的事件？两个相互独立事件同时发生的概率与每个事件概率之间有什么关系？
4. 设在一次试验中，某事件发生的概率是 p，如何计算在 n 次独立重复试验中这个事件恰好发生 k 次的概率？
5. 随机变量的数学期望和方差的直观意义是什么？总结一下常见的几个分布的期望和方差的计算公式.
6. 什么是正态分布和正态分布的 3σ 原则？

Ⅲ 巩固与提高

1. 判断下列命题的真假：

（1）对任意两个事件 A，B 都有 $P(A\cap B)=P(A)\times P(B)$；

(2) 如果事件 A 发生，事件 B 一定发生，则 $P(A\cap B)=P(B)$；

(3) 已知在一次试验中 $P(A)=0.1$，那么在 3 次独立重复试验中 A 恰好发生 2 次的概率是 $C_3^2\cdot(0.1)^{3-2}\cdot(0.9)^2=3\times0.1\times0.81=0.243$；

(4) 抛掷一枚硬币 100 次，则正面向上出现的次数超过 40 次.

2. 填空：

(1) 抛掷 3 枚硬币，至少出现一个正面的概率等于______；

(2) 已知在一次试验中，$P(A)=0.6$，那么在 3 次独立重复试验中，事件 A 恰好发生一次的概率是______；

(3) 甲、乙两人在相同的条件下进行投篮，甲投中的概率是 0.8，乙投中的概率是 0.9，两人各投篮一次，恰有一人投中的概率是______；

(4) 某篮球运动员投篮一次投中的概率是 0.8，他投篮 4 次恰好投中 3 次的概率等于______.

3. 盒中有 5 个白球和 4 个黑球，从中任意取出 3 个，设 X 表示其中黑球的个数，试求 X 的分布列.

4. 生产一种零件，甲车间的合格率是 96%，乙车间的合格率是 97%，从它们生产的零件中各抽取一件，都抽到合格品的概率是多少？

5. 有发芽率分别是 0.9 和 0.7 的两批种子，在两批种子中各取一粒，求下列事件发生的概率：

(1) 两粒种子都发芽；

(2) 恰有一粒种子发芽；

(3) 至少有一粒种子发芽.

6. 某篮球运动员进行 5 次定点投篮，已知每次投中的概率均为 0.4. 求 5 次投篮恰有 2 次投中的概率.

7. 某大型国有企业为 10 000 名员工定制工作服，设员工的身高（单位：cm）服从正态分布 $N(172,5^2)$. 试估计适宜身高在 167～177 cm 范围内员工穿的服装大约要定制多少套？

Ⅳ 自测与评估

1. 设随机变量 X 的概率分布为 $P(X=k)=\dfrac{m}{k}\ (k=1,2,3,4)$：

(1) 确定常数 m 的值；

(2) 写出 X 的分布列；

(3) 计算 $P(1<X<4)$.

2. 某种大炮击中目标的概率是 0.3，以多少门这样的大炮同时射击一次，就可使击中目标的概率超过 95%.

3. 甲、乙两运动员进行乒乓球单打比赛，根据以往比赛的胜负情况知道，每一局甲胜的

概率 0.6，乙胜的概率为 0.4．如果比赛可采用“三局两胜”或“五局三胜”两种制度，求在哪种比赛制度下，甲胜的概率较大？

4. 有 10 道判断对错的测验题，一个人随意猜答，他答对不少于 6 道题的概率是多少？

5. 甲、乙两人进行投篮比赛，每人投 3 次，投中的次数分别记作 X，Y．它们的分布列如下：

X	0	1	2	3
P	0.1	0.6	0.2	0.1

Y	0	1	2	3
P	0.3	0.3	0.2	0.2

试比较甲、乙两人哪位投篮较稳定？

6. 一批白炽灯泡的光通量服从 $N(209,\ 6.5^2)$，如果有这样一批灯泡 10 000 个，试估计光通量在下列范围内的灯泡的个数：

(1) $209-2\times6.5\sim209+2\times6.5$；

(2) $209-3\times6.5\sim209+3\times6.5$．

数 学 文 化

阅读与欣赏

关于“玛丽莲问题”的争论

“玛丽莲问题”是某电视台娱乐节目“决策”中提出的一系列问题．其中最著名的是“Behind Monty Hall's Doors”．问题如下：台上有三个关闭的门，一个后边有汽车，其余两个后边是山羊．主持人让参加者任意选择其中一个门，然后她打开其余两个门中的一个，参加者看到的是山羊．这时，她让参加者可以重选，也就是说参加者可以换选另一个剩下的门．那么，参加者应该换还是不换？玛丽莲的答案是应该换，但是很多读者不同意．玛丽莲在下一期专栏给出一个表格说明她的道理，但反对声更多更大了．在几千封读者来信中，反对者达九成．其中有全国健康机构的统计学家、国防情报中心的副主任，甚至著名的美籍匈牙利数学家保罗·埃尔德希(Paul Erdos)也是反对者之一．

那么到底应该换还是不换呢？我们先来考虑3个人通过抽签分一张演唱会票这一问题．3个人按排定的顺序从分别写有“有票”“无票”“无票”的3个纸团中各抽一个来决定谁能得到演唱会票，每个人得票的概率是多少呢？

对第一个人来说，从3个纸团中任取一个，得票的概率为$\frac{1}{3}$．为了求出第二个人得票的概率，我们来分析一下前两个人抽取纸团的情况．从3个纸团中先后抽出2个，可以看成从3个元素中取出2个进行排列，它的种数是A_3^2，而其中第二个人得票的情况有A_2^1种，因此第一个人未得票，而第二个人得票的概率为

$$\frac{A_2^1}{A_3^2}=\frac{1}{3}.$$

通过类似的分析，可知第三个人得票的概率为

$$\frac{A_2^2}{A_3^3}=\frac{1}{3},$$

由此看出，不管抽取纸团的次序如何，每个人得到演唱会票的概率都是$\frac{1}{3}$．

“玛丽莲问题”也可以看作是一个分票问题，如果用写有“汽车”的纸团代表汽车，写有“山羊”的纸团代表山羊．三次决定可以看作三个人各抽取一个纸团，第一个决定是参加者作出的，相当于第一个人抽取一个纸团，得到汽车的概率是$\frac{1}{3}$，第二个决定是主持人作出的，如果主持人是随机作出的决定，那么他得到汽车的概率也是$\frac{1}{3}$．第三个决定仍由参加者作出，如果他换选剩下的另一个门，就相当于第三个人抽取纸团，由上面分票的结果知得到汽车的概率仍是$\frac{1}{3}$，这样换与不换得到汽车的概率相等．

但是，由前面知道主持人的决定并不是随机的，她知道哪一个门后面有汽车．所以主持人打开的门后面总是山羊．仍从分票来考虑，这相当于第一个人抽取纸团后，第二个人抽取的总是写有“山羊”的纸团．即第一个人获得汽车的概率是$\frac{1}{3}$，第二个人获得汽车的概率是0，第三个人获得汽车的概率为$1-\frac{1}{3}-0=\frac{2}{3}$，所以换选后得到汽车的概率为$\frac{2}{3}$，参加者应该改变自己的选择．

附 录

1. 超几何分布的数学期望

我们先来证明一个公式

$$mC_n^m=nC_{n-1}^{m-1}. \quad ①$$

证明：根据组合数公式，

$$\text{左式}=m\times\frac{n!}{m!\,(n-m)!}=\frac{n!}{(m-1)!\,(n-m)!},$$

$$\text{右式}=n\times\frac{(n-1)!}{(m-1)!\,(n-m)!}=\frac{n!}{(m-1)!\,(n-m)!},$$

左式=右式，公式得证.

由公式①立刻可以得到

$$C_N^M=\frac{N}{M}C_{N-1}^{M-1}. \quad ②$$

下面我们来求超几何分布的期望，设随机变量 X 服从参数为 N，M，n 的超几何分布，则 X 的分布列为

$$P(X=m)=\frac{C_n^mC_{N-n}^{M-m}}{C_N^M}\ (m=0,1,\cdots,l,\ l\text{ 为 }n\text{ 和 }M\text{ 中较小的一个}).$$

同二项分布类比，我们猜想它的期望可能是 $n\cdot\frac{M}{N}$. 由数学期望的定义式得

$$E(X)=\sum_{m=0}^{l}m\cdot P(X=m)=\sum_{m=0}^{l}m\cdot\frac{C_n^mC_{N-n}^{M-m}}{C_N^M}=\sum_{m=1}^{l}\frac{m\cdot C_n^m\cdot C_{N-n}^{M-m}}{C_N^M}$$

$$=\sum_{m=1}^{l}n\cdot C_{n-1}^{m-1}\cdot\frac{M}{N}\cdot\frac{C_{N-n}^{M-m}}{C_{N-1}^{M-1}} \quad (\text{由公式 ①②})$$

$$=n\cdot\frac{M}{N}\cdot\sum_{m=1}^{l}C_{n-1}^{m-1}\cdot\frac{C_{N-n}^{M-m}}{C_{N-1}^{M-1}} \quad (\text{令 } m-1=i)$$

$$=n\cdot\frac{M}{N}\cdot\sum_{i=0}^{l-1}C_{n-1}^{i}\cdot\frac{C_{(N-1)-(n-1)}^{(M-1)-i}}{C_{N-1}^{M-1}}$$

$$=n\cdot\frac{M}{N}.$$

上式中 $C_{n-1}^{i}\cdot\frac{C_{(N-1)-(n-1)}^{(M-1)-i}}{C_{N-1}^{M-1}}$ 可以看作 $N-1$ 件产品中有 $n-1$ 件次品，从中任取 $M-1$件（$M\leqslant N$），其中恰有 i 件次品的概率，所以对于 $i=0,1,\cdots,l-1$ 求和得 1.

2. 二项分布的方差

设随机变量 $X\sim B(n,p)$，则 X 的分布列是

$$P(X=k)=C_n^kp^kq^{n-k},k=0,1,\cdots,n.$$

$$P(X=0)+P(X=1)+\cdots+P(X=n)=\sum_{k=0}^{n}C_n^kp^kq^{n-k}=1,$$

$$E(X)=\sum_{k=0}^{n}k\cdot C_n^kp^kq^{n-k}=np.$$

由方差的公式得

$$
\begin{aligned}
D(X) &= \sum_{k=0}^{n}[x_k - E(X)]^2 \cdot p_k = \sum_{k=0}^{n}(k-np)^2 \cdot \mathrm{C}_n^k p^k q^{n-k} \\
&= \sum_{k=0}^{n}k^2 \cdot \mathrm{C}_n^k p^k q^{n-k} + \sum_{k=0}^{n} -2npk \cdot \mathrm{C}_n^k p^k q^{n-k} + \sum_{k=0}^{n} n^2 p^2 \cdot \mathrm{C}_n^k p^k q^{n-k} \\
&= \sum_{k=0}^{n}(k-1)\cdot k \cdot \mathrm{C}_n^k p^k q^{n-k} + \sum_{k=0}^{n} k \cdot \mathrm{C}_n^k p^k q^{n-k} - 2np\sum_{k=0}^{n} k \cdot \mathrm{C}_n^k p^k q^{n-k} \\
&\quad + n^2 p^2 \sum_{k=0}^{n}\mathrm{C}_n^k p^k q^{n-k} \\
&= \sum_{k=1}^{n}(k-1)\cdot k \cdot \mathrm{C}_n^k p^k q^{n-k} + np - 2np \cdot np + n^2 p^2 \\
&= n\sum_{k=2}^{n}(k-1)\cdot \mathrm{C}_{n-1}^{k-1} p^k q^{n-k} + np - n^2 p^2 \qquad \text{（公式①）} \\
&= n(n-1)p^2 \sum_{k=2}^{n}\mathrm{C}_{n-2}^{k-2} p^{k-2} q^{(n-2)-(k-2)} + np - n^2 p^2 \quad \text{（令 } k-2=i\text{）} \\
&= n(n-1)p^2 \sum_{i=0}^{n-2}\mathrm{C}_{n-2}^{i} p^{i} q^{(n-2)-i} + np - n^2 p^2 \\
&= n(n-1)p^2 + np - n^2 p^2 \\
&= npq.
\end{aligned}
$$

第三章　统计案例

	晕　机	不晕机	合计
男　性	24	31	55
女　性	8	26	34
合　计	32	57	89

我们在必修模块数学3中，学习过一些统计知识，接触到诸如随机抽样、用样本估计总体、线性回归分析等方法．实际上，统计知识的应用远不止于此．在一个逐步实现现代化的社会里，统计信息将越来越多，这促使我们去学习对一些统计信息进行分析、推断的本领．这里，先举两个例子．

有人对一老年烟民劝道："你快戒烟吧，否则一定会患慢性气管炎的．"他的话有没有道理？老年人患慢性气管炎与吸烟习惯有没有关系？

从一些纪实电视片或推理小说中常常看到这样的情节：刑警在案发现场仔细地搜寻罪犯的脚印，其理由之一是，我们可以根据一个人的脚印长度来预测他的身高，上述理由的根据是什么呢？

独立性检验和回归分析是解决这些问题的统计方法．它们在国民经济和日常生活的很多方面有着广泛的应用．线性回归分析的部分内容同学们在必修模块数学3中有所接触，本章将进一步讨论线性回归分析的一些问题，并介绍非线性回归分析的初步知识．

本章分2节，每节讨论一种统计方法．各节的编写特点是，把一个个的案例直接呈现在同学们面前，通过探究案例，解决问题，使同学们了解这两种统计方法的基本思想、解题步骤及其初步应用．

独立性检验和回归分析只是丰富多彩的统计世界的两个部分，我们欢迎同学们进入这个世界，并希望你们能喜爱这个世界．

3.1 独立性检验

例1 为了探究患慢性气管炎是否与吸烟有关，调查了339名50岁以上的人，调查结果如下表所示：

	患慢性气管炎	未患慢性气管炎	合计
吸　　烟	43	162	205
不 吸 烟	13	121	134
合　　计	56	283	339

试问：50岁以上的人患慢性气管炎与吸烟习惯有关吗？

分析 例1中给出的表称为2×2列联表，意思是问题中要考虑50岁以上的人的两种状态：是否吸烟，是否患慢性气管炎；每种状态又分两种情况：吸烟，不吸烟以及患慢性气管炎，未患慢性气管炎．表中排成两行两列的数据是调查得来的结果，希望根据这4个数据来检验上述两种状态是否有关．这一检验问题就称为2×2列联表的独立性检验．

下面进一步分析独立性检验的含义．

为了把问题讨论清楚，并便于向一般情况推广，我们用字母来代替2×2列联表中的事件和数据，得到一张用字母来表示的2×2列联表，如下表所示：

	患慢性气管炎(B)	未患慢性气管炎(B)	合计
吸　烟(A)	n_{11}	n_{12}	n_{1+}
不吸烟(A)	n_{21}	n_{22}	n_{2+}
合　　计	n_{+1}	n_{+2}	n

注

右表中 $n_{1+}=n_{11}+n_{12}$，$n_{2+}=n_{21}+n_{22}$，$n_{+1}=n_{11}+n_{21}$，$n_{+2}=n_{12}+n_{22}$，$n=n_{11}+n_{12}+n_{21}+n_{22}$．

(1) 首先，当吸烟(A)与患慢性气管炎(B)无关时，用概率方法进行推理，看看会出现什么结果．

上面的话的意思是指事件A与B独立，这时应该有

$$P(AB)=P(A)P(B).$$

成立．我们用字母H_0表示上式，即

$$H_0: P(AB)=P(A)P(B),$$

并称之为统计假设，当H_0成立时，下面的三个式子也都成立：

$$P(AB)=P(A)P(B),\ P(AB)=P(A)P(B),\ P(AB)=P(A)P(B).$$

根据概率的统计定义，上面提到的众多事件的概率都可用相应的频率来估计．例如，$P(AB)$的估计为$\frac{n_{11}}{n}$，$P(A)$的估计为$\frac{n_{1+}}{n}$，$P(B)$的估计为$\frac{n_{+1}}{n}$……

于是$\frac{n_{11}}{n}$与$\frac{n_{1+}}{n}\cdot\frac{n_{+1}}{n}$应该很接近，$\frac{n_{12}}{n}$与$\frac{n_{1+}}{n}\cdot\frac{n_{+2}}{n}$应该很接近……

或者说，

$$\left(\frac{n_{11}}{n}-\frac{n_{1+}}{n}\cdot\frac{n_{+1}}{n}\right)^2,\ \left(\frac{n_{12}}{n}-\frac{n_{1+}}{n}\cdot\frac{n_{+2}}{n}\right)^2,\ \left(\frac{n_{21}}{n}-\frac{n_{2+}}{n}\cdot\frac{n_{+1}}{n}\right)^2,\ \left(\frac{n_{22}}{n}-\frac{n_{2+}}{n}\cdot\frac{n_{+2}}{n}\right)^2$$

应该比较小，从而

$$\frac{\left(\frac{n_{11}}{n}-\frac{n_{1+}}{n}\cdot\frac{n_{+1}}{n}\right)^2}{\frac{n_{1+}}{n}\cdot\frac{n_{+1}}{n}}+\frac{\left(\frac{n_{12}}{n}-\frac{n_{1+}}{n}\cdot\frac{n_{+2}}{n}\right)^2}{\frac{n_{1+}}{n}\cdot\frac{n_{+2}}{n}}+\frac{\left(\frac{n_{21}}{n}-\frac{n_{2+}}{n}\cdot\frac{n_{+1}}{n}\right)^2}{\frac{n_{2+}}{n}\cdot\frac{n_{+1}}{n}}+\frac{\left(\frac{n_{22}}{n}-\frac{n_{2+}}{n}\cdot\frac{n_{+2}}{n}\right)^2}{\frac{n_{2+}}{n}\cdot\frac{n_{+2}}{n}} \quad ①$$

也应该比较小.

（2）上面的表达式①就是统计中非常有用的χ^2（读作“卡方”）统计量，它可以化简为

$$\chi^2=\frac{n(n_{11}n_{22}-n_{12}n_{21})^2}{n_{1+}n_{2+}n_{+1}n_{+2}}. \quad ②$$

用它的大小可以决定是否拒绝原来的统计假设H_0. 如果算出的χ^2值较大，就拒绝H_0，也就是拒绝“事件A与B无关”，从而就认为它们是有关的了.

（3）两个临界值：3.841与6.635.

经过对χ^2统计量分布的研究，已经得到了两个临界值：3.841与6.635. 当根据具体的数据算出的$\chi^2>3.841$时，有95%的把握说事件A与B有关；当$\chi^2>6.635$时，有99%的把握说事件A与B有关. 当$\chi^2\leqslant 3.841$时，认为事件A与B是无关的.

对于例1，最理想的解决办法是向所有50岁以上的人做调查，然后对得到的数据进行统计处理，但这花费的代价太大，实际上是行不通的. 339个人相对于全体50岁以上的人，只是一个小部分. 回忆一下数学3(必修)中学过的总体和样本的关系，当用样本平均数、样本标准差去估计总体相应的数字特征时，由于抽样的随机性，结果并不唯一. 现在情况类似，我们用部分对全体作推断，推断可能正确，也可能错误. 例如我们知道，不少中老年烟民的身体很好，没有患慢性气管炎；而又有很多从不吸烟的中老年人体质很差，患有慢性气管炎. 如果抽取的339个调查对象中很多人来自上述两个群体，试想会得出什么结论吧. 我们有95%(或99%)的把握说事件A与B有关，是指推断犯错误的可能性为5%(或1%)，这也常常说成是“以95%(或99%)的概率”，其含义是一样的.

解：由公式②，

$$\chi^2=\frac{339\times(43\times121-162\times13)^2}{205\times134\times56\times283}$$
$$\approx 7.469.$$

因为 7.469>6.635，所以我们有99%的把握说：50岁以上的人患慢性气管炎与吸烟习惯有关.

例2 对196个接受心脏搭桥手术的病人和196个接受血管清障手术的病人进行了3年的跟踪研究，调查他们是否又发作过心脏病，调查结果如下表所示：

	又发作过心脏病	未发作心脏病	合计
心脏搭桥手术	39	157	196
血管清障手术	29	167	196
合　计	68	324	392

试根据上述数据比较这两种手术对病人又发作心脏病的影响有没有差别.

解：这是一个 2×2 列联表的独立性检验问题，由公式②，

$$\chi^2=\frac{392\times(39\times167-157\times29)^2}{196\times196\times68\times324}\approx1.780.$$

因为 1.780<3.841，我们没有理由说“心脏搭桥手术”与“又发作过心脏病”有关，可以认为病人又发作心脏病与否跟他做过何种手术无关.

这里我们再提醒一句，上述结论是对所有做过心脏搭桥手术或血管清障手术的病人而言的，绝不要误以为只对 392 个跟踪研究对象成立.

例 3　某大型企业人力资源部为了研究企业员工工作积极性和对待企业改革态度的关系，随机抽取了 189 名员工进行调查，所得数据如下表所示：

	积极支持企业改革	不太赞成企业改革	合计
工作积极	54	40	94
工作一般	32	63	95
合　计	86	103	189

对于人力资源部的研究项目，根据上述数据能得出什么结论？

解：由公式②，

$$\chi^2=\frac{189\times(54\times63-40\times32)^2}{94\times95\times86\times103}\approx10.759.$$

因为 10.759>6.635，所以有 99%的把握说：员工“工作积极”与“积极支持企业改革”是有关的，可以认为企业的全体员工对待企业改革的态度与其工作积极性是有关的.

例 4　在一次恶劣气候的飞行航程中调查男女乘客在机上晕机的情况如下表所示，据此资料你是否认为在恶劣气候飞行中男性比女性更容易晕机？

	晕　机	不晕机	合计
男　性	24	31	55
女　性	8	26	34
合　计	32	57	89

解：由公式②，

$$\chi^2=\frac{89\times(24\times26-31\times8)^2}{55\times34\times32\times57}\approx3.689.$$

因为 3.689<3.841，我们没有理由说晕机与否跟男女性别有关，尽管这次航班中男

性晕机的比例$\left(\frac{24}{55}\right)$比女性晕机的比例$\left(\frac{8}{34}\right)$高，但我们不能认为在恶劣气候飞行中男性比女性更容易晕机.

注意　使用 χ^2 统计量作 2×2 列联表的独立性检验时，要求表中的 4 个数据都要大于 5，为此，在选取样本的容量时一定要注意这一点. 本例的 4 个数据 24，31，8，26 都大于 5，是满足这一要求的.

探索与研究

2×2 列联表中的 4 个数据为什么都要大于 5 呢？有兴趣的同学可以通过查阅一些专业统计书籍来了解相关知识.

例 5　打鼾不仅影响别人休息，而且可能与患某种疾病有关. 下表是一次调查所得的数据，试问：每一晚都打鼾与患心脏病有关吗？

	患心脏病	未患心脏病	合计
每一晚都打鼾	30	224	254
不 打 鼾	24	1 355	1 379
合　　计	54	1 579	1 633

解：由公式②，

$$\chi^2=\frac{1\,633\times(30\times1\,355-224\times24)^2}{1\,379\times254\times54\times1\,579}\approx68.033.$$

因为 $68.033>6.635$，所以有 99% 的把握说，每一晚都打鼾与患心脏病有关.

注意　本例和例 1 类似，我们所说“每一晚都打鼾与患心脏病有关”或“患慢性气管炎与吸烟习惯有关”指的是统计上的关系，不要误以为这里是因果关系. 具体到某一个每一晚都打鼾的人，并不能说他患心脏病. 其实从 2×2 列联表中也可看出，每一晚都打鼾的人群中，患心脏病的概率也只有$\frac{30}{254}$，稍微超过十分之一，至于他患不患心脏病，应该由医学检查来确定，已经不是统计学的事了.

本节通过对 5 个例子的探究来讨论两个事件是否独立，在 2×2 列联表的独立性检验中，我们选用了 χ^2 统计量，可以用它的取值大小来推断独立性是否成立．独立性检验在生物统计、医学统计等学科的应用很广泛，在处理调查社会问题得到的数据时，也常常使用独立性检验．

习题 3-1 A

1. 在 500 个人身上试验某种血清预防感冒的作用，把一年中的记录与另外 500 个未用血清的人作比较，结果如下：

	未感冒	感　冒	合计
处　理	252	248	500
未处理	224	276	500
合　计	476	524	1 000

问该种血清能否起到预防感冒的作用？

2. 考察小麦种子经灭菌与否跟发生黑穗病的关系，经试验观察，得到数据如下表所示：

	种子灭菌	种子未灭菌	合计
黑穗病	26	184	210
无黑穗病	50	200	250
合　　计	76	384	460

试按照原试验目的作统计分析推断．

3. 调查者通过询问 72 名男女大学生在购买食品时是否看营养说明，得到的数据如下表所示：

	看营养说明	不看营养说明	合计
男大学生	28	8	36
女大学生	16	20	36
合　　计	44	28	72

问大学生的性别和是否看营养说明之间有没有关系？

4. 在研究某种新措施对猪白痢的防治效果问题时，得到了以下数据：

	存活数	死亡数	合计
新措施	132	18	150
对　照	114	36	150
合　计	246	54	300

试问新措施对防治猪白痢是否有效?

习题 3-1 B

全班同学请分成一些小组，每组 4～5 名同学，在老师的指导下，开展一次简单的调查活动，并对调查结果进行统计分析.

重新阅读本节例 4 和习题 3-1A 的第 3 题，实际上在我们的周围，男女同学对很多问题的看法可能有差别，也可能没有差别，可以用独立性检验的方法作出统计推断.

要求每个小组设计一个同学们较关心的只有两种答案的问题，如“你打算报考理工类高校吗”“你对美容的态度”“你喜欢上外语课吗”等，通过询问同学取得数据，作独立性检验，并分析得到的结果，最后写出一份简明的调查报告.

注意决定样本容量时必须保证所取得的 4 个数据都大于 5.

3.2 回归分析

例 1 研究某灌溉渠道水的流速 Y 与水深 x 之间的关系，测得一组数据如下：

水深 x/m	1.40	1.50	1.60	1.70	1.80	1.90	2.00	2.10
流速 Y/(m·s^{-1})	1.70	1.79	1.88	1.95	2.03	2.10	2.16	2.21

(1) 求 Y 对 x 的回归直线方程；

(2) 预测水深为 1.95 m 时水的流速是多少？

分析 从散点图可以直观地看出变量 x 与 Y 之间有无线性相关关系，为此把这 8 对数据描绘在平面直角坐标系中，得到平面上 8 个点，如图 3-1 所示.

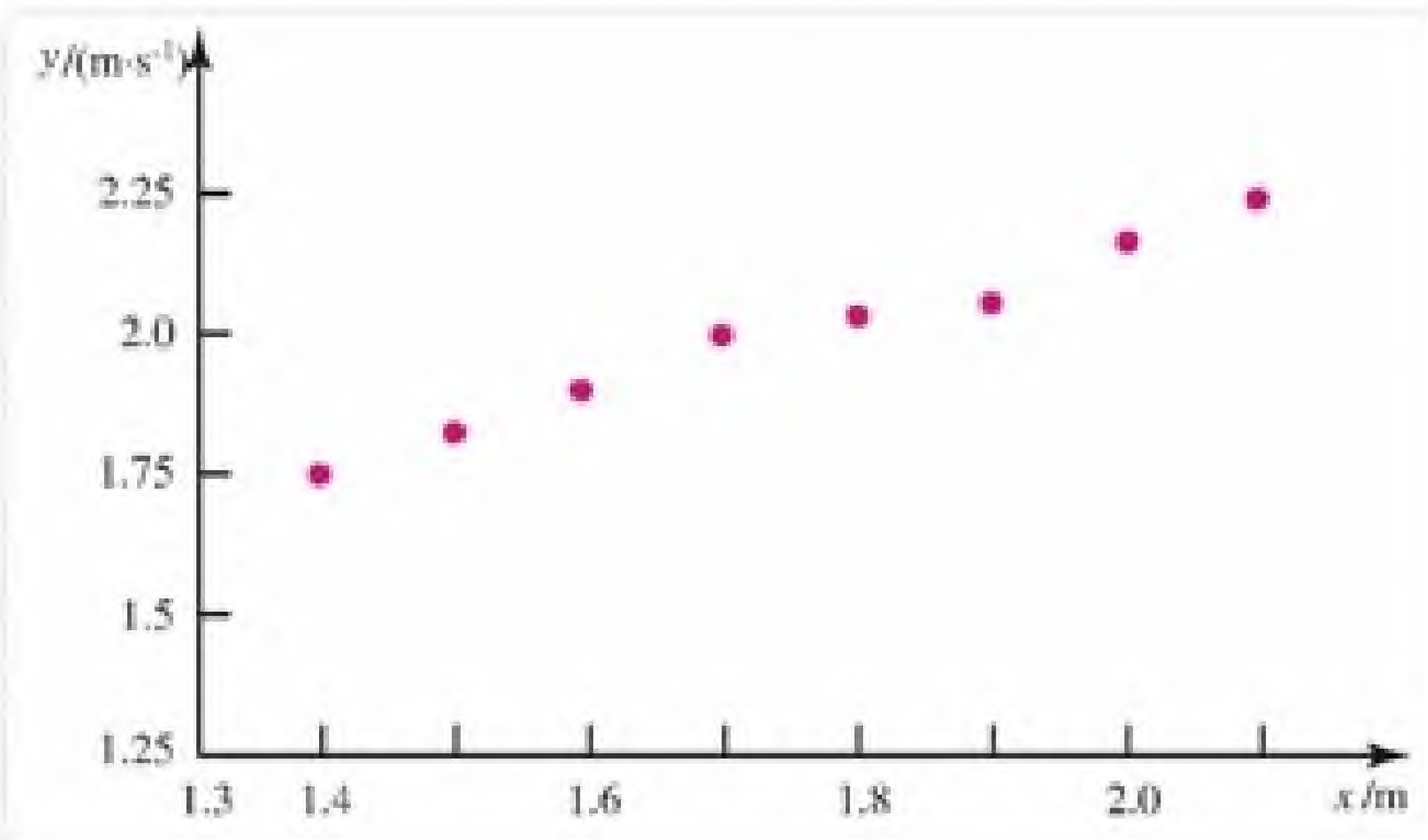


图 3-1

由图 3-1 容易看出，x 与 Y 之间有近似的线性相关关系，或者说，可以用一个回归直线方程

$$\hat{y}=a+bx$$

来反映这种关系，这些是我们在必修模块数学 3 中学过的知识.

进一步观察这 8 个点，容易发现它们并不是“严格地”在一条直线上. 对于某个 x_i，由上式能确定一个 $\hat{y}_i=a+bx_i$，一般地说，由于测量流速可能存在误差，或者受某些随机因素的影响，或者上面的回归直线方程本身就不够精确，$\hat{y}_i$ 与测得的数据 y_i 很可能不相等，即

$$y_i=\hat{y}_i+\varepsilon_i \ (i=1, 2, \cdots, 8),$$

其中 ε_i 是随机误差项. 于是，就有

$$y_i=a+bx_i+\varepsilon_i \ (i=1, 2, \cdots, 8),$$

这就是本题的线性模型.

从上述线性模型出发，我们可以求出 a 与回归系数 b 的估计值 $\hat{a}$，$\hat{b}$，使得全部误差

ε_1，ε_2，…，ε_8 的平方和达到最小，当然，这是一种很好的估计. 最后得到的求 $\hat{a}$，$\hat{b}$ 的数学公式为

$$\hat{b}=\frac{\sum_{i=1}^{8}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{8}(x_i-\bar{x})^2},$$

$$\hat{a}=\bar{y}-\hat{b}\bar{x}.$$

探索与研究

我们研究在一般情况下（已知 n 对数据 (x_1, y_1)，(x_2, y_2)，…，(x_n, y_n) 即 n 个点）如何推导求 $\hat{a}$ 与 $\hat{b}$ 的公式.

随机误差

$$\varepsilon_i=y_i-a-bx_i,\ i=1,\ 2,\ \cdots,\ n.$$

假如把这些随机误差直接相加作为总的误差，是很不合理的，因为它们有正有负，相加起来可能抵消一部分. 为了不使误差之和正负抵消，我们设全部误差的平方和为

$$Q=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2,$$

用 Q 的大小来度量总的误差大小. Q 是 a，b 的二元函数，可记作 $Q(a, b)$，用下面的配方法可以求出 $Q(a, b)$ 达到最小值时 a，b 所取的值.

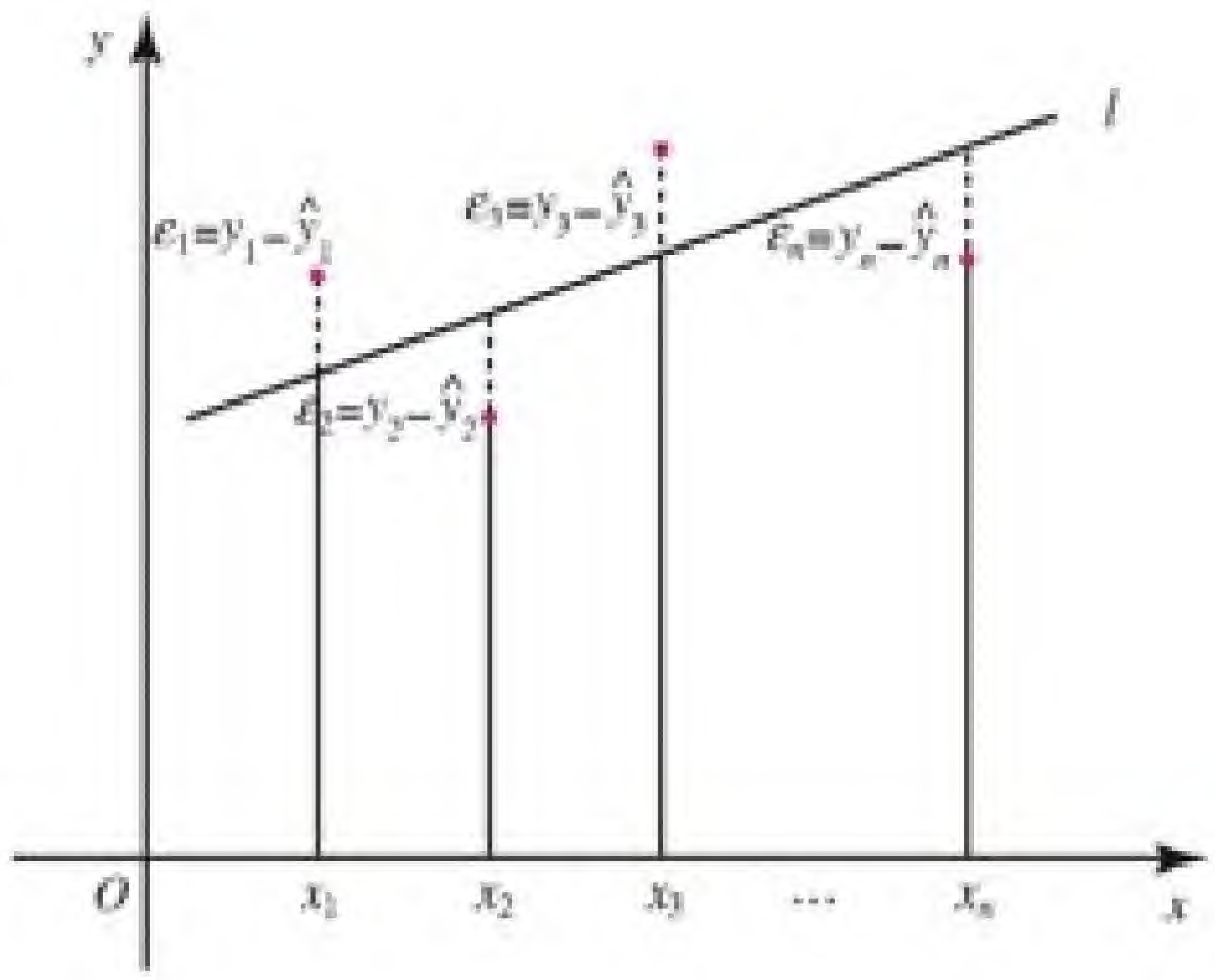


图 3-2

记 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$，我们有

$$\begin{aligned}
Q(a, b)&=\sum_{i=1}^{n}\{(y_i-\bar{y})+[\bar{y}-(a+b\bar{x})]-b(x_i-\bar{x})\}^2\\
&=\sum_{i=1}^{n}(y_i-\bar{y})^2+n[\bar{y}-(a+b\bar{x})]^2+b^2\sum_{i=1}^{n}(x_i-\bar{x})^2+2[\bar{y}-(a+b\bar{x})]\cdot\sum_{i=1}^{n}(y_i-\bar{y})\\
&\quad-2b[\bar{y}-(a+b\bar{x})]\cdot\sum_{i=1}^{n}(x_i-\bar{x})-2b\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\\
&=\sum_{i=1}^{n}(y_i-\bar{y})^2+n[\bar{y}-(a+b\bar{x})]^2+b^2\sum_{i=1}^{n}(x_i-\bar{x})^2-2b\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\\
&=\sum_{i=1}^{n}(y_i-\bar{y})^2+n[\bar{y}-(a+b\bar{x})]^2+\sum_{i=1}^{n}(x_i-\bar{x})^2\left[b^2-2b\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]
\end{aligned}$$

$$=\sum_{i=1}^{n}(y_i-\bar{y})^2+n[\bar{y}-(a+b\bar{x})]^2+\sum_{i=1}^{n}(x_i-\bar{x})^2\left[b-\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\right]^2$$

$$-\frac{\left[\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}.$$

经过略嫌冗长的推导，这些推导的目的是把 a，b “配” 到含有平方项的底的中间去．对于 n 对数据来说，x_1，x_2，…，x_n 一般不会相等(否则这 n 对数据已经在一条平行于 y 轴的直线上了，再求回归直线已失去意义)，所以

$$\sum_{i=1}^{n}(x_i-\bar{x})^2\neq 0.$$

观察上面最后的表达式，其中 y_i，$\bar{y}$，n，$\bar{x}$，x_i 都是已知数，含 a，b 的两项是非负数，当且仅当它们等于 0 时，$Q(a,b)$取最小值．这就是说，当

$$\hat{b}=\frac{\sum\limits_{i=1}^{n}(x_i-x)(y_i-y)}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2},$$

$$\hat{a}=y-\hat{b}x$$

时 $Q(a,b)$达到最小值．上面的 $\hat{b}$ 也可以进一步推导成

$$\hat{b}=\frac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2}.$$

至此我们推导出了求 a 与回归系数 b 的数学公式．

解：(1) 由上面的分析，可采用列表的方法计算 a 与回归系数 b．

序号	x	y	x^2	xy
1	1.40	1.70	1.96	2.380
2	1.50	1.79	2.25	2.685
3	1.60	1.88	2.56	3.008
4	1.70	1.95	2.89	3.315
5	1.80	2.03	3.24	3.654
6	1.90	2.10	3.61	3.990
7	2.00	2.16	4.00	4.320
8	2.10	2.21	4.41	4.641
合计	14.00	15.82	24.92	27.993

于是，

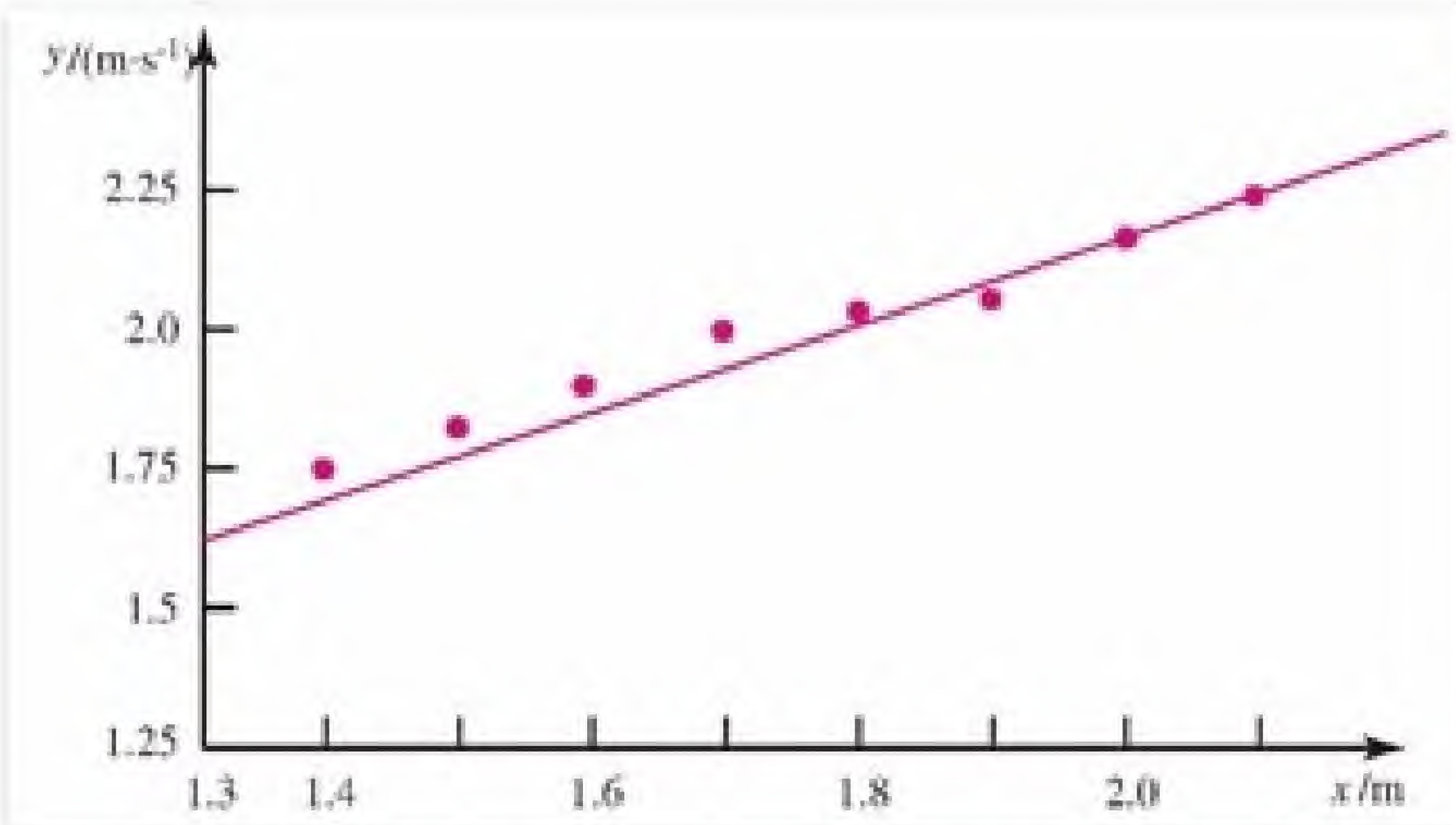
$$\bar{x}=\frac{1}{8}\times 14.00=1.75,\ \bar{y}=\frac{1}{8}\times 15.82=1.977\,5.$$

$$\hat{b}=\frac{27.993-8\times 1.75\times 1.977\,5}{24.92-8\times 1.75^2}=\frac{11}{15}\approx 0.733.$$

$$\hat{a}=1.977\,5-\frac{11}{15}\times 1.75\approx 0.694.$$

Y 对 x 的回归直线方程为

$$\hat{y}=\hat{a}+\hat{b}x=0.694+0.733x.$$



回归系数 $\hat{b}=0.733$ 的意思是，在此灌溉渠道中，水深每增加 0.1 m，水的流速平均增加 0.073 m/s（本例数据是以 0.1 m 为水深间隔测得的），$\hat{a}=0.694$ 可以解释为水的流速中不受水深影响的部分.

（2）由(1)中求出的回归直线方程，把 $x=1.95$ 代入，易得

$$\hat{y}=0.694+0.733\times 1.95\approx 2.12(\text{m/s}).$$

计算结果表明，当水深为 1.95 m 时可以预测渠水的流速约为 2.12 m/s.

例 2　为了了解某地母亲身高 x 与女儿身高 Y 的相关关系，随机测得 10 对母女的身高如下表所示：

母亲身高 x/cm	159	160	160	163	159	154	159	158	159	157
女儿身高 Y/cm	158	159	160	161	161	155	162	157	162	156

试对 x 与 Y 进行一元线性回归分析，并预测当母亲身高为 161 cm 时女儿的身高为多少？

分析　把这 10 对数据画出散点图如图 3-3 所示，其中点(159，162)表示两对母女的身高数据，可以看出，x 与 Y 之间有近似的线性相关关系.

散点图能帮助我们寻找线性相关关系，既直观又方便．只需要用一张坐标纸，把已知的成对数据标在直角坐标系中便可得到散点图．即使没有坐标纸，改用普通白纸也可以，因为我们并不要求把点标得十分准确，只要能看出这些点大致分布在某条直线附近就可以了.

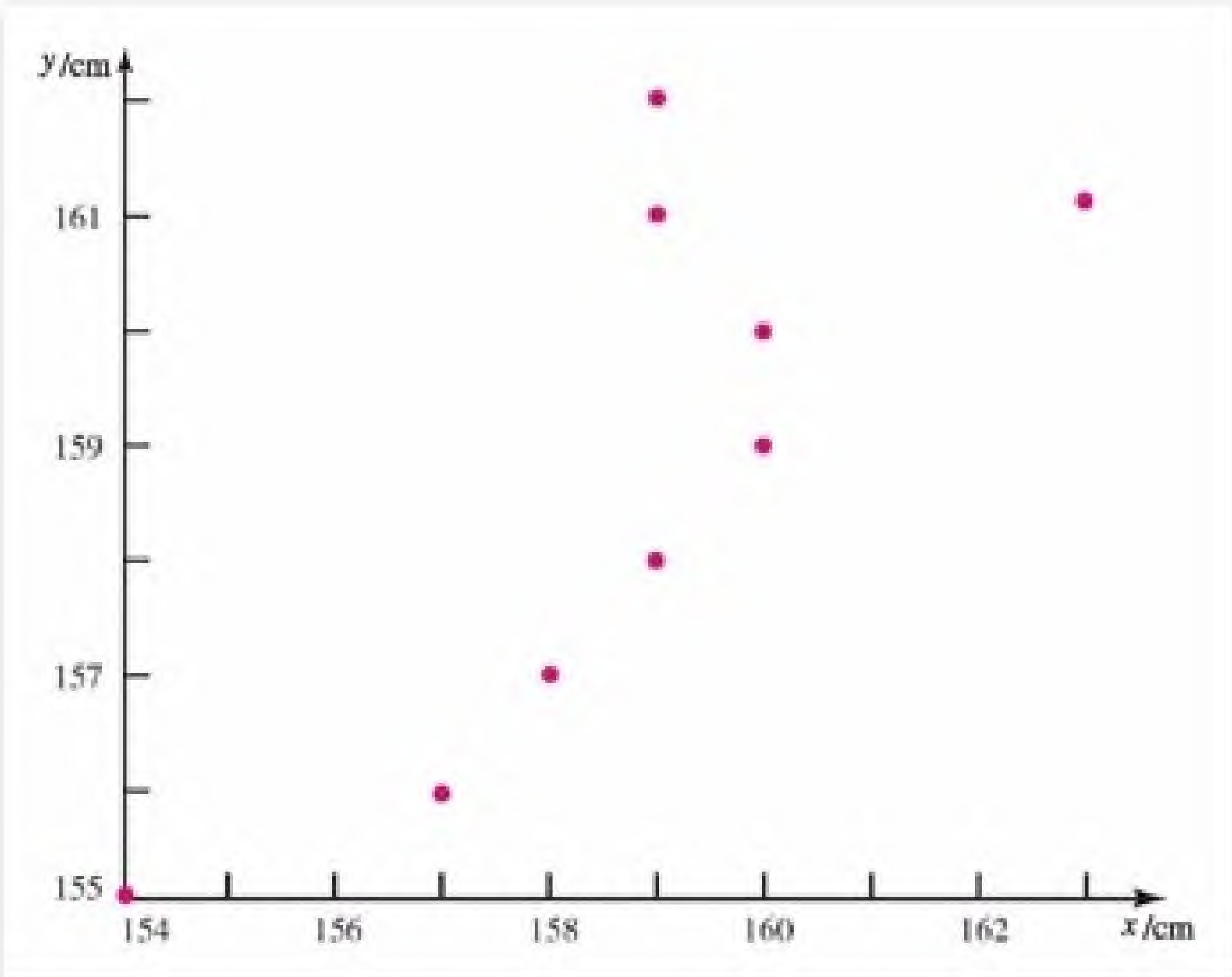


图 3-3

麻烦在于有时很难说这些点是不是分布在某条直线附近，如图 3-4 的两个散点图，都很难下判断. 右边那个图中散布着的点更像在一条曲线附近.

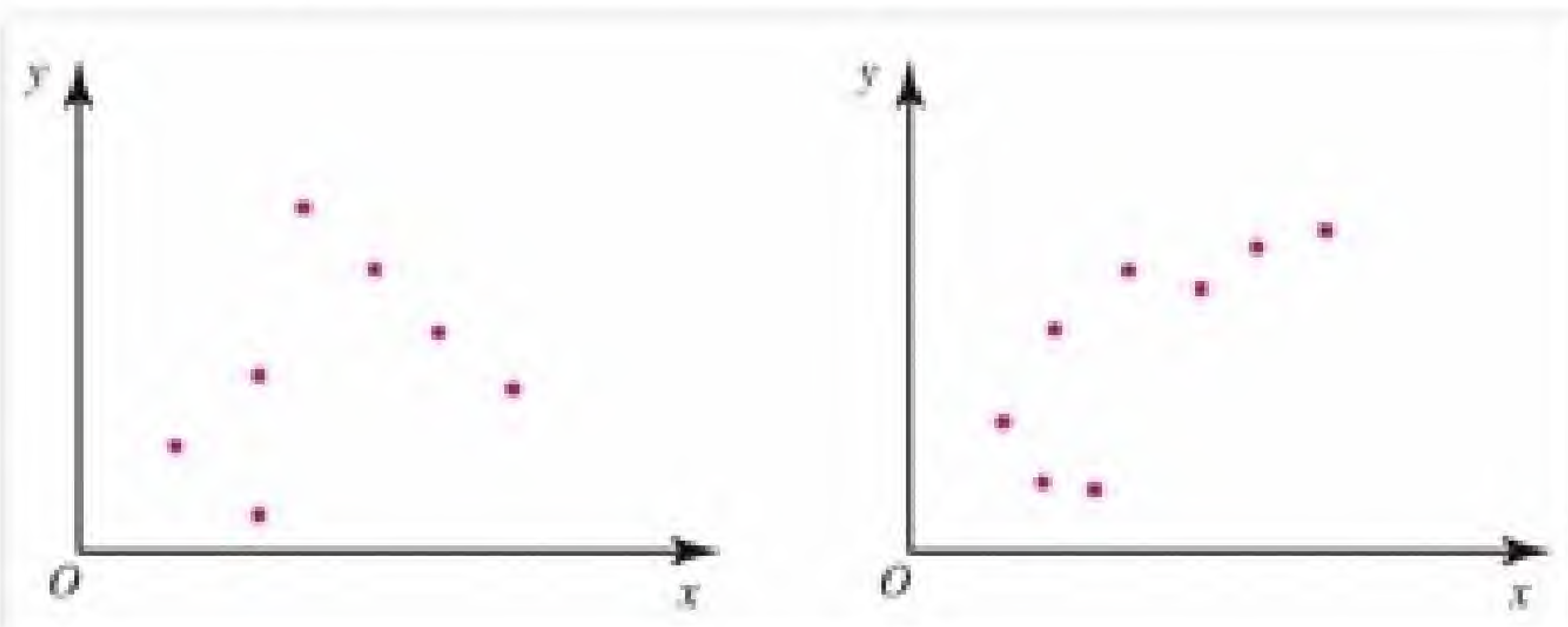


图 3-4

此外，假如不考虑散点图，按照例 1 给出的计算 a 与回归系数 b 的公式，我们可以根据一组成对数据，求出一个回归直线方程. 但它能不能反映这组成对数据的变化规律？如不能，这又有多少实际意义呢？

为了解决上述问题，我们有必要对 x 与 Y 作线性相关性检验，简称相关性检验.

对于变量 x 与 Y 随机抽取到的 n 对数据 (x_1, y_1)，(x_2, y_2)，…，(x_n, y_n)，检验统计量是样本相关系数

$$r=\frac{\sum_{i=1}^{n}(x_i-x)(y_i-y)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}=\frac{\sum_{i=1}^{n}x_iy_i-nxy}{\sqrt{(\sum_{i=1}^{n}x_i^2-n\bar{x}^2)(\sum_{i=1}^{n}y_i^2-n\bar{y}^2)}}.$$

r 具有以下性质：$|r|\leqslant 1$，并且 $|r|$ 越接近 1，线性相关程度越强；$|r|$ 越接近 0，线性相关程度越弱.

检验的步骤如下：

（1）作统计假设：x 与 Y 不具有线性相关关系.

（2）根据小概率 0.05 与 $n-2$ 在附表中查出 r 的一个临界值 $r_{0.05}$.

（3）根据样本相关系数计算公式算出 r 的值.

（4）作统计推断. 如果 $|r|>r_{0.05}$，表明有 95%的把握认为 x 与 Y 之间具有线性相关关系.

> 注
>
> 这里可回顾 3.1 节独立性检验的例 1，那里给出的 3.841 是 χ^2 统计量的一个临界值.

如果 $|r|\leqslant r_{0.05}$，我们没有理由拒绝原来的假设. 这时寻找回归直线方程是毫无意义的.

解：由以上分析，先对 x 与 Y 作相关性检验.

（1）作统计假设：x 与 Y 不具有线性相关关系.

（2）由小概率 0.05 与 $n-2=8$ 在附表中查得

$$r_{0.05}=0.632.$$

（3）$\bar{x}=\frac{1}{10}(159+160+\cdots+157)=158.8$，

$\bar{y}=\frac{1}{10}(158+159+\cdots+156)=159.1$，

$\sum\limits_{i=1}^{10}x_i^2-10\bar{x}^2=(159^2+160^2+\cdots+157^2)-10\times158.8^2$
$=47.6$，

$\sum\limits_{i=1}^{10}x_iy_i-10\bar{x}\bar{y}=(159\times158+160\times159+\cdots+157\times156)$
$-10\times158.8\times159.1$
$=37.2$，

$\sum\limits_{i=1}^{10}y_i^2-10\bar{y}^2=(158^2+159^2+\cdots+156^2)-10\times159.1^2$
$=56.9$，

所以

$$r=\frac{37.2}{\sqrt{47.6\times56.9}}\approx0.71.$$

（4）$|r|=0.71>0.632$，即

$$|r|>r_{0.05}，$$

从而有 95%的把握认为 x 与 Y 之间具有线性相关关系，去求回归直线方程是有意义的.

回归系数

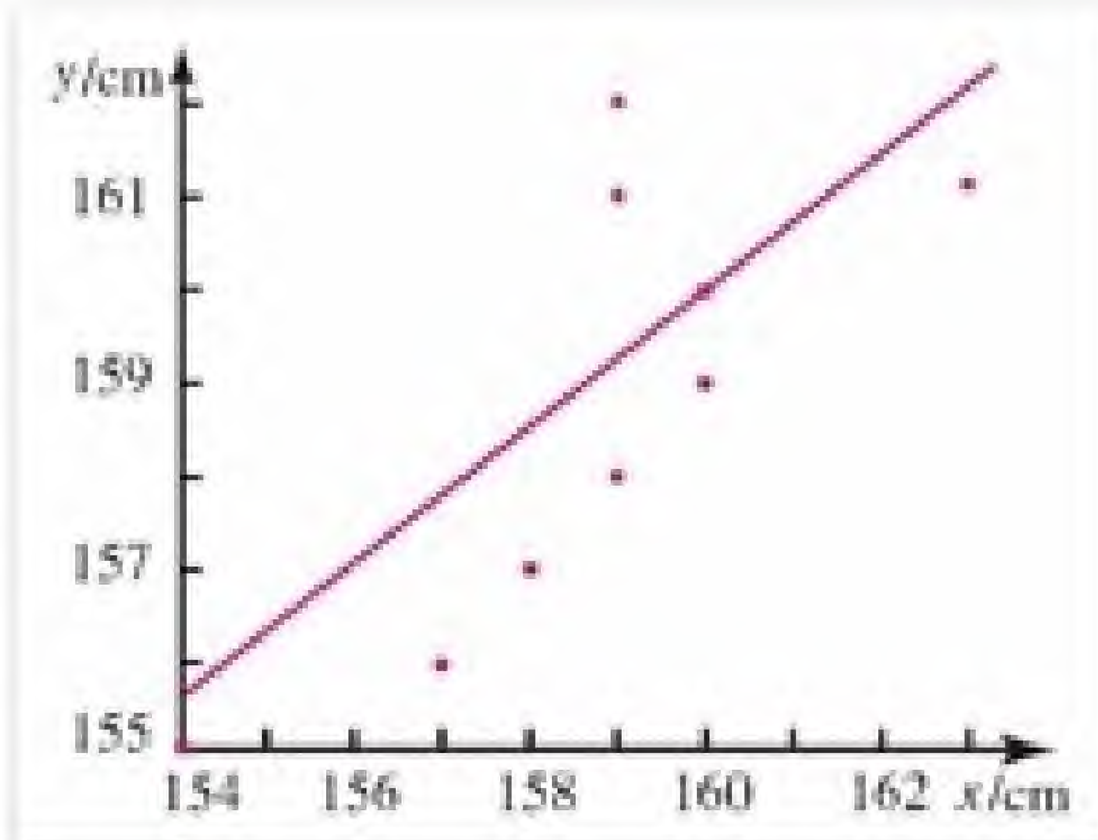
$$\hat{b}=\frac{37.2}{47.6}\approx0.782\approx0.78，$$

$$\hat{a}=159.1-0.782\times158.8\approx34.92.$$

所以 Y 对 x 的回归直线方程是

$$\hat{y}=34.92+0.78x.$$

回归系数 0.78 反映出当母亲身高每增加 1 cm 时女儿身高平均增加 0.78 cm，



$\hat{a}=34.92$可以解释为女儿身高不受母亲身高变化影响的部分.

当 $x=161$ 时

$$\hat{y}=34.92+0.78\times 161=160.5.$$

这就是说当母亲身高为 161 cm 时女儿的身高大致也接近 161 cm.

例 3 某市居民 1996～2003 年货币收入 x 与购买商品支出 Y 的统计资料如下表所示：

年　份	1996	1997	1998	1999	2000	2001	2002	2003
货币收入 x/亿元	36	37	38	40	42	44	47	50
购买商品支出 Y/亿元	30.0	31.0	32.0	33.2	34.8	36.5	39.0	41.6

试对 x 与 Y 的关系进行相关性检验，如 x 与 Y 具有线性相关关系，求出 Y 对 x 的回归直线方程（结果保留 3 位小数）.

分析 由上述例 1、例 2 可见，回归系数与样本相关系数的计算工作相当繁重，有条件的同学应尽量借助计算器来完成. 现在国内生产的供高中生使用的科学函数计算器普遍都设置了进入回归计算的专用按键，使用起来十分方便. 我们希望同学们在掌握了回归分析的基本思想方法后，只要有条件，应该使用计算器等现代技术手段来处理数据. 下面主要介绍如何使用函数型计算器解决问题.

解：本例数据的散点图如 3-5 所示，题目要求我们对 x 与 Y 作相关性检验.

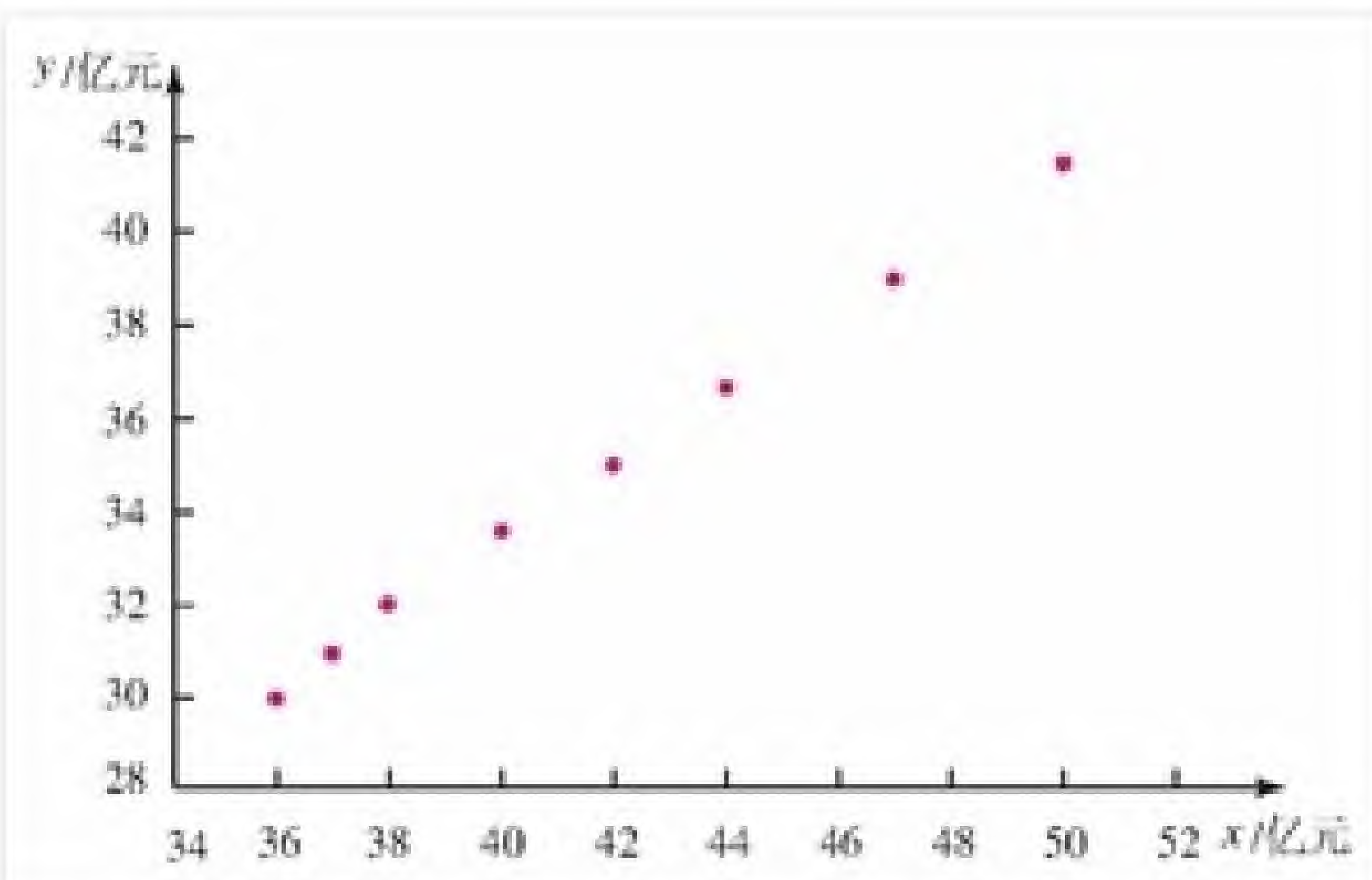


图 3-5

（1）作统计假设：x 与 Y 不具有线性相关关系.

（2）由小概率 0.05 与 $n-2=6$ 在附表中查得

$$r_{0.05}=0.707.$$

（3）使用函数型计算器进行计算.

按键

$\boxed{\text{MODE}}$ $\boxed{3}$ $\boxed{1}$（进入线性回归计算状态）

$\boxed{\text{SHIFT}}$ $\boxed{\text{CLR}}$ $\boxed{1}$ $\boxed{=}$（将计算器存储器设置成初始状态）

36 $\boxed{,}$ 30.0 $\boxed{\text{DT}}$ 37 $\boxed{,}$ 31.0 $\boxed{\text{DT}}$ 38 $\boxed{,}$ 32.0 $\boxed{\text{DT}}$ 40 $\boxed{,}$ 33.2 $\boxed{\text{DT}}$

42 $\boxed{,}$ 34.8 $\boxed{\text{DT}}$ 44 $\boxed{,}$ 36.5 $\boxed{\text{DT}}$ 47 $\boxed{,}$ 39.0 $\boxed{\text{DT}}$ 50 $\boxed{,}$ 41.6 $\boxed{\text{DT}}$

继续按下表按键：

按键	显示结果
$\boxed{\text{SHIFT}}$ $\boxed{\text{S-VAR}}$ $\boxed{\blacktriangleright}$ $\boxed{\blacktriangleright}$ $\boxed{3}$ $\boxed{=}$	0.999246378
$\boxed{\text{SHIFT}}$ $\boxed{\text{S-VAR}}$ $\boxed{\blacktriangleright}$ $\boxed{\blacktriangleright}$ $\boxed{1}$ $\boxed{=}$	0.851152737
$\boxed{\text{SHIFT}}$ $\boxed{\text{S-VAR}}$ $\boxed{\blacktriangleright}$ $\boxed{\blacktriangleright}$ $\boxed{2}$ $\boxed{=}$	0.812247838

（4）$|r|=0.999>0.707$，即

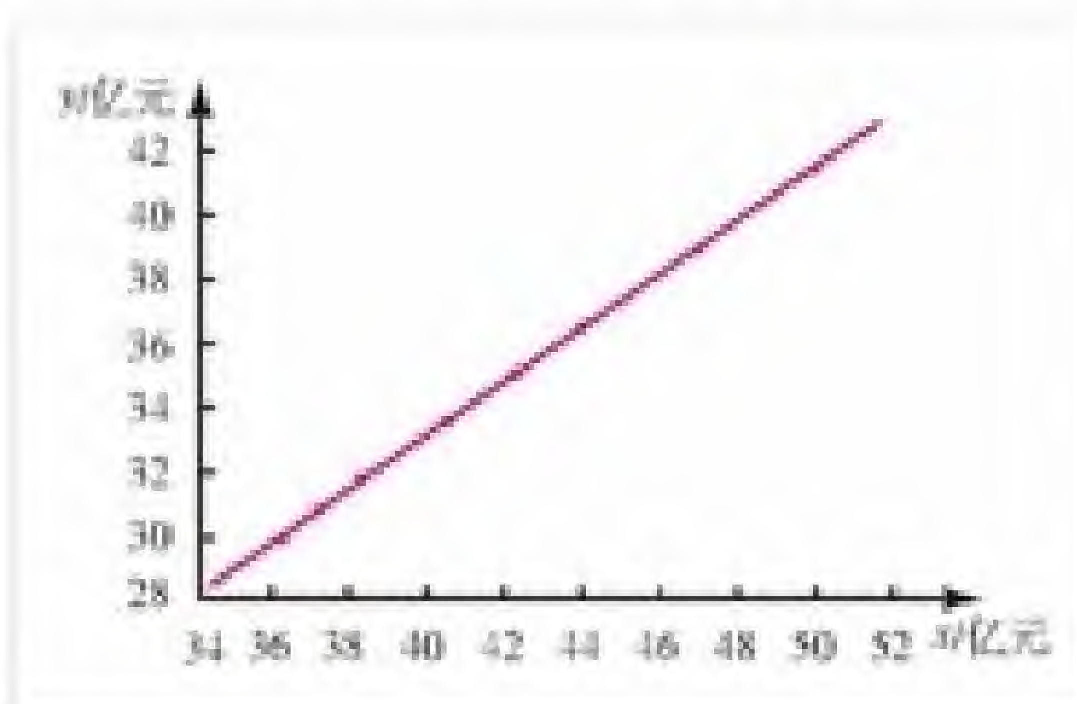
$$|r|>r_{0.05}.$$

从而有95%的把握认为x与Y之间具有线性相关关系，求回归直线方程是有意义的.

（5）上表中最后两行给出的A与B就是$\hat{a}$与回归系数$\hat{b}$，所以Y对x的回归直线方程是

$$\hat{y}=0.851+0.812x,$$

其中回归系数0.812是指居民每增加1亿元的货币收入，大约会有0.812亿元用于购买商品，$\hat{a}=0.851$可以解释为购买商品支出中不受货币收入变化影响的部分，反映了最基本的购买需求.



例4　在彩色显影中，由经验知：形成染料光学密度Y与析出银的光学密度x由公式

$$y=Ae^{\frac{b}{x}}\ (b<0)$$

表示．现测得试验数据如下：

x_i	0.05	0.06	0.25	0.31	0.07	0.10
y_i	0.10	0.14	1.00	1.12	0.23	0.37
x_i	0.38	0.43	0.14	0.20	0.47	
y_i	1.19	1.25	0.59	0.79	1.29	

试求Y对x的回归方程.

分析　本例与前面三个例子不同，是非线性回归分析问题．由于题目已给出了所要求的曲线$y=Ae^{\frac{b}{x}}$类型，只要通过已知的11对样本数据，把A与b确定下来，就找到了描述x与Y相关关系（注意，这里不是线性相关关系）的一条函数曲线.

在此我们特别指出，确定性关系(如公式、函数关系等)和相关关系之间并没有一条不可逾越的鸿沟．由于有实验误差、测量误差等存在，变量之间的确定性关系往往通过相关关系表现出来．反过来，在有些问题中，可以通过研究相关关系来深入了解变量变化的内在规律，从而找到它们的确定性关系．

由于我们已经基本掌握了线性回归分析的思想方法，所以解决本例的思路是通过适当的变量置换，把非线性回归方程化成线性回归方程，然后再套用前三例的解题步骤．

解：根据以上分析，由题给的经验公式

$$y=Ae^{\frac{b}{x}},$$

两边取自然对数，便得

$$\ln y=\ln A+\frac{b}{x}.$$

与线性回归直线方程相对照，只要取

$$u=\frac{1}{x},\ v=\ln y,\ a=\ln A,$$

就有

$$v=a+bu,$$

这是 V 对 u 的线性回归直线方程．对此我们已掌握了一套相关性检验、求 a 与回归系数 b 的方法．

题给数据经变量置换 $u=\frac{1}{x}$，$v=\ln y$ 变成如下表所示的数据：

u_i	20.000	16.667	4.000	3.226	14.286	10.000
v_i	−2.303	−1.966	0	0.113	−1.470	−0.994
u_i	2.632	2.326	7.143	5.000	2.128	
v_i	0.174	0.223	−0.528	−0.236	0.255	

(1) 作统计假设：u 与 V 不具有线性相关关系．

(2) 由小概率 0.05 与 $n-2=9$ 在附表中查得

$$r_{0.05}=0.602.$$

(3) 使用函数型计算器进行计算．

按键

$\boxed{\text{MODE}}$ $\boxed{3}$ $\boxed{1}$ (进入线性回归计算状态)

$\boxed{\text{SHIFT}}$ $\boxed{\text{CLR}}$ $\boxed{1}$ $\boxed{=}$ (将计算器存储器设置成初始状态)

20.000 $\boxed{,}$ −2.303 $\boxed{\text{DT}}$ 16.667 $\boxed{,}$ −1.966 $\boxed{\text{DT}}$ 4.000 $\boxed{,}$ 0 $\boxed{\text{DT}}$ 3.226 $\boxed{,}$ 0.113 $\boxed{\text{DT}}$
14.286 $\boxed{,}$ −1.470 $\boxed{\text{DT}}$ 10.000 $\boxed{,}$ −0.994 $\boxed{\text{DT}}$ 2.632 $\boxed{,}$ 0.174 $\boxed{\text{DT}}$ 2.326 $\boxed{,}$ 0.223 $\boxed{\text{DT}}$
7.143 $\boxed{,}$ −0.528 $\boxed{\text{DT}}$ 5.000 $\boxed{,}$ −0.236 $\boxed{\text{DT}}$ 2.128 $\boxed{,}$ 0.255 $\boxed{\text{DT}}$

继续按下表按键：

按键	显示结果
SHIFT S-VAR ▶ ▶ 3 =	-0.998283392
SHIFT S-VAR ▶ ▶ 1 =	0.547673504
SHIFT S-VAR ▶ ▶ 2 =	-0.145940972

（4）$|r|=0.998>0.602$，即

$$|r|>r_{0.05},$$

从而有 95%的把握认为 u 与 V 之间具有线性相关关系，求 V 对 u 的回归直线方程有意义.

（5）由上表立刻可得

$$\hat{v}=0.548-0.146u.$$

把 u 与 v 换回原来的变量 x 与 y，即

$$u=\frac{1}{x},\ v=\ln y,$$

可得

$$\ln\hat{y}=0.548-\frac{0.146}{x},$$

即

$$\begin{aligned}\hat{y}&=e^{0.548-\frac{0.146}{x}}=e^{0.548}\cdot e^{-\frac{0.146}{x}}\\&=1.73e^{-\frac{0.146}{x}}.\end{aligned}$$

这就是 Y 对 x 的回归曲线方程，试验点及回归曲线的图形如图 3-6 所示.

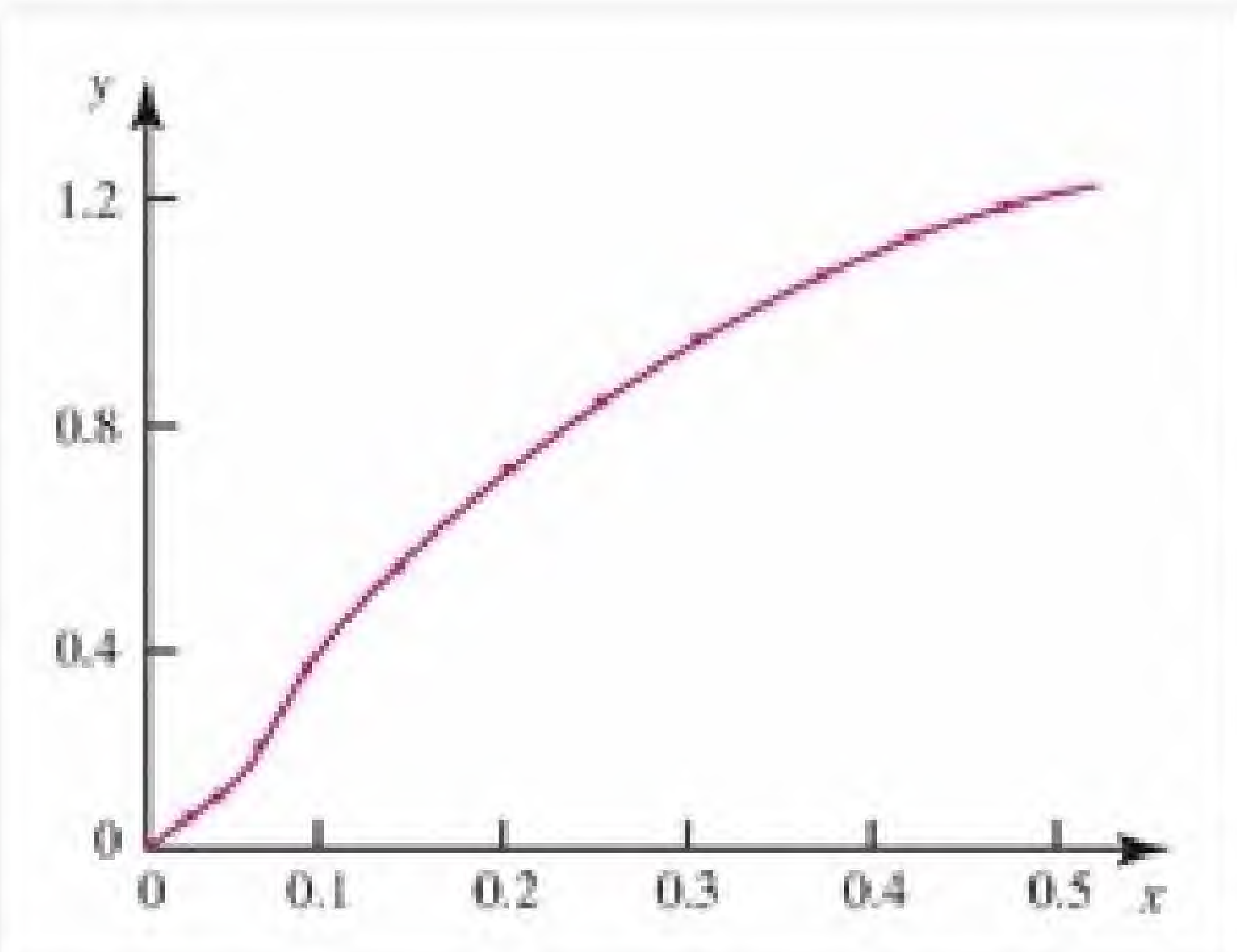


图 3-6

在实践中，两个变量间的关系可能是线性相关，但在很多情况下呈现出一种“曲线关系”，例如在生物医学的毒理学试验中，毒物的浓度与动物死亡率之间的关系、污染物与其跟污染源距离之间的关系等都是“曲线关系”．由例 4 可以看出，研究非线性回归问题要比研究线性回归问题复杂得多.

注意 非线性回归问题有时并不给出经验公式．这时我们可以画出已知数据的散点图，把它与必修模块数学 1 中学过的各种函数（幂函数、指数函数、对数函数、logistic模型的“S”形曲线函数等）图象作比较，挑选一种跟这些散点拟合得最好的函数，然后像例 4 那样，采用适当的变量置换，把问题化为线性回归分析问题，使之得到解决．

我们在必修模块数学 3 中已经初步讨论过一元线性回归分析问题．由于回归分析在生产实际和日常生活中的应用很广泛，本节通过 4 个例子的探究，使同学们进一步了解回归分析（包括非线性回归分析）的基本思想、方法及其初步应用，具体做题时应尽量借助计算器．

习题 3-2 A

1. 在一段时间内，某种商品价格 x（万元）和需求量 Y(t)之间的一组数据为

价　格 x：	1.4	1.6	1.8	2	2.2
需求量 Y：	12	10	7	5	3

(1) 画出散点图；

(2) 求出 Y 对 x 的回归直线方程，并在 (1) 的散点图中画出它的图象；

(3) 如价格定为 1.9 万元，预测需求量大约是多少（精确到 0.01 t)?

2. 弹簧长度 Y(cm) 随所挂物体质量 x(g) 不同而变化的情况如下：

物体质量 x：	5	10	15	20	25	30
弹簧长度 Y：	7.25	8.12	8.95	9.90	10.96	11.80

(1) 画出散点图；

(2) 进行相关性检验；

(3) 求 Y 对 x 的回归直线方程；

(4) 预测所挂物体质量为 27 g 时的弹簧长度（精确到 0.01 cm).

3. 某农场对单位面积化肥用量 x(kg) 和水稻相应产量 Y(kg) 的关系作了统计，得到数据如下：

x：	15	20	25	30	35	40	45
Y：	330	345	365	405	445	450	455

(1) 进行相关性检验；

(2) 如果 x 与 Y 之间具有线性相关关系，求出回归直线方程，并预测当单位面积化肥用量为 32 kg 时水稻的产量大约是多少（精确到 0.01 kg).

4. 某种书每册的成本费 Y（元）与印刷册数 x（千册）有关，经统计得到数据如下：

x：	1	2	3	5	10	20	30	50	100	200
Y：	10.15	5.52	4.08	2.85	2.11	1.62	1.41	1.30	1.21	1.15

检验每册书的成本费 Y 与印刷册数的倒数 $\frac{1}{x}$ 之间是否具有线性相关关系，如有，求出 Y 对 x 的回归方程.

5. 电容器充电后，电压达到 100 V，然后开始放电，测得时间 t(s)时的电压 U(V)如下所示：

t：	0	1	2	3	4	5	6	7	8	9	10
U：	100	75	55	40	30	20	15	10	10	5	5

试对变量 t 与 U 作回归分析并求出 U 对 t 的回归方程.

提示：电压 U 随时间 t 按指数规律衰减，有

$$U=Ae^{bt}\ (b<0).$$

习题 3-2 B

我们知道，刑警如果能在案发现场提取到罪犯的脚印，那将获得一条重要的破案线索. 其原因之一是人类的脚掌长度和身高存在着相关关系，可以根据一个人的脚掌长度来预测他的身高……

我们还知道，在统计史上，很早就有人收集过人们的身高、前臂长度等数据，试图寻找这些数据之间的规律……

在上述两个小故事的启发下，全班同学请分成一些小组，每组 4～5 名同学，在老师的指导下，开展一次数学建模活动，来亲自体验回归分析的思想方法，提高自己的实践能力.

数学建模的题目是：收集一些周围人们的脚掌长度、前臂长度中的一个数据及其身高，作为两个变量画散点图、进行相关性检验，如果这两个变量之间具有线性相关关系，就求出回归直线方程. 另选一人的这两个变量的数据，作一次预测，并分析预测结果.

最后按小组写出数学建模报告. 报告要求过程清晰，结论明确，有关数学论述准确. 以下两个问题需要注意：

(1) 如量脚掌长度不方便，可改量脚印的长度；

(2) 数据尽量取得分散一些.

本章小结

Ⅰ 知识结构

- 统计案例
  - 独立性检验
  - 回归分析

Ⅱ 思考与交流

1. χ^2 统计量在 2×2 列联表的独立性检验中起了什么作用？

2. 本章的回归分析内容与必修模块数学 3 的回归分析内容相比，你有哪些新的体会．

Ⅲ 巩固与提高

1. 学习过独立性检验后，总结一下你对统计推断的认识．

2. 学习过回归分析后，总结一下你对在统计方法中应用现代计算技术的认识．

Ⅳ 自测与评估

1. 在调查的 480 名男人中 38 名患有色盲，520 名女人中 6 名患有色盲，试检验人的性别与患色盲与否是否独立？

2. 某车间为了规定工时定额，需要确定加工零件所花费的时间，为此作了 10 次试验，得到数据如下：

零件数 x/个：	10	20	30	40	50	60	70	80	90	100
加工时间 Y/min：	62	68	75	81	89	95	102	108	115	122

试对上述变量作回归分析，你认为工时定额定多少比较合理？

数学文化　　阅读与欣赏

"回归"一词的由来

在3.2节的例2中，通过对10对母女身高的探究，我们求出了女儿身高对母亲身高的回归直线方程，从中可以看出，一般说来，母亲身材较高者，其女儿的身材也较高．但为什么要称这样的直线方程为回归直线方程呢？总觉得与"回归"两字有点沾不上边．

原来问题并非如此简单．"回归"作为统计学的一个术语，最早来自英国人类学家兼统计学家高尔顿(Galton)的"普用回归定律"概念．他的学生、现代统计学的奠基者之一皮尔逊(Pearson)曾收集了1 078对父亲及其一个成年儿子的身高数据．若用x表示父亲身高，Y表示儿子身高，单位用英寸(1英寸约为2.54 cm)．高尔顿把这1 078对数据标在直角坐标纸上，发现散点图大致呈直线状．也就是说，总的趋势是父亲身材较高者，其儿子的身材也较高，这和上面母女身高的情况类似，也和我们的常识一致．

经过高尔顿对数据的深入分析，发现这1 078个x_i的平均值是68英寸，1 078个y_i的平均值是69英寸．这就是说，子代身高平均增加了1英寸．人们自然会想，若父亲身高为x，那么他儿子的身高平均来说大致是$x+1$．但高尔顿的研究结果却与上述想法大相径庭．他发现，当父亲身高为72英寸时，他们的儿子平均身高仅为71英寸，并没有达到预期的73英寸．若父亲身高只有64英寸，他们的儿子平均身高为67英寸，竟比预期的65英寸高了2英寸．这一事实反映出子代身高有向平均值69英寸"回归"的倾向．

高尔顿对此的解释是：大自然具有一种约束力，使人类身高的分布在一定时期内相对稳定而不产生两极分化，这就是所谓的"回归"效应．通过这一例子，高尔顿引入了"回归"一词．用这1 078对数据，可以算出x与Y之间存在线性关系：

$$\hat{y}=33.73+0.516x,$$

它在几何上代表一条直线，人们通常就把它称为回归直线方程了．

随着时代的发展，回归分析的应用越来越广泛．事实上，对于大多数实际问题来说，两个有着线性相关关系的变量并不具备回归效应，"回归"分析有点名不符实．但是，由于"回归"一词沿用已久，今天实在没有必要对3.2节讨论的这种统计分析方法另外取一个名字了．

附　表

相关性检验的临界值表

$n-2$	小概率		$n-2$	小概率	
	0.05	0.01		0.05	0.01
1	0.997	1.000	16	0.468	0.590
2	0.950	0.990	17	0.456	0.575
3	0.878	0.959	18	0.444	0.561
4	0.811	0.917	19	0.433	0.549
5	0.754	0.874	20	0.423	0.537
6	0.707	0.834	21	0.413	0.526
7	0.666	0.798	22	0.404	0.515
8	0.632	0.765	23	0.396	0.505
9	0.602	0.735	24	0.388	0.496
10	0.576	0.708	25	0.381	0.487
11	0.553	0.684	26	0.374	0.478
12	0.532	0.661	27	0.364	0.470
13	0.514	0.641	28	0.361	0.463
14	0.497	0.623	29	0.355	0.456
15	0.482	0.606	30	0.349	0.449

注：表中的 n 为数据的对数.

附 录

部分中英文词汇对照表

排列	arrangement
排列数	number of arrangement
全排列	total arrangement
阶乘	factorial
组合	combination
组合数	number of combination
二项式定理	binomial theorem
二项式系数	binomial coefficient
二项展开式	binomial expansion
概率	probability
相互独立事件	mutually independent events
独立重复试验	independent repeated trials
随机变量	random variable
离散型随机变量	discrete random variable
概率分布	probability distribution
分布列	distribution series
数学期望	mathematical expectation
方差	variance
标准差	standard deviation
正态分布	normal distribution
标准正态分布	standard normal distribution
线性回归	linear regression
统计	statistics
相关	correlation
相关系数	correlation coefficient
散点图	scatter diagram
回归直线	regression straight line
线性回归分析	linear regression analysis
独立性检验	test of independence

后 记

根据教育部制订的普通高中各学科课程标准(实验)，人民教育出版社课程教材研究所编写的各学科普通高中课程标准实验教科书，得到了诸多教育界前辈和各学科专家学者的热情帮助和大力支持. 在各学科教科书终于同课程改革实验区的师生见面时，我们特别感谢担任教科书总顾问的丁石孙、许嘉璐、叶至善、顾明远、吕型伟、王梓坤、梁衡、金冲及、白春礼、陶西平同志，感谢担任教科书编写指导委员会主任委员的柳斌同志和编写指导委员会委员的江蓝生、李吉林、杨焕明、顾泠沅、袁行霈等同志.

本套高中数学实验教科书(B版)的总指导为丁尔陞教授. 从教材立项、编写、送审到进入实验区实验的过程中，在丁尔陞、孙瑞清、江守礼、房艮孙、王殿军等专家教授的指导下，经过实验研究组全体成员的努力，基本上完成了“课标”中各模块的编写任务，并通过了教育部的审查.

山东、辽宁等实验区的教研员和教师在实验过程中，对教材编写的指导思想、教材内容的科学性、基础性、选择性以及是否易教、易学等诸方面，进行了审视和检验，提出了许多的宝贵意见，并针对教材和教学写出了大量的论文. 我们在总结实验的基础上，逐年对教材进行认真的修改，使教材不断的完善. 现在所取得的成果，是实验研究组全体成员、编者，实验区的省、市、县各级教学研究员及广大数学教师集体智慧的结晶.

各实验区参加教材审读、研讨及修改的主要成员有：

韩际清、常传洪、尹玉柱、秦玉波、祝广文、尚凡青、杨长智、田明泉、邵丽云、于世章、李明照、胡廷国、张颖、张成钢、李学生、朱强、窦同明、姜传祯、韩淑勤、王宗武、黄武吕.

刘莉、宋明新、高锦、赵文莲、王孝宇、周善富、胡文亮、孙家逊、舒凤杰、齐力、林文波、教丽、刘鑫、李凤、金盈、潘戈、高钧、魏明智、刘波、崔贺、李忠、关玲、郝军、郭艳霞、董晖、赵光千、王晓声、王文、姚琳.

在此，特向参与、帮助、支持这套教科书编写的专家、学者和教师深表谢意.

我们还要感谢实验区的教育行政和教研部门，以及使用本套教材的师生们.

让我们与一切关心这套教材建设的朋友们，共同携起手来，为建设一套具有中国特色的高中数学教材而努力.

我们的联系方式如下：

电话：010-58758523 010-58758532　电子邮件：longzw@pep.com.cn

人民教育出版社　课程教材研究所

中学数学教材实验研究组